INVESTIGATING A ROLE FOR COHESIN IN GENOME ORGANISATION AND GENE REGULATION IN POST-MITOTIC ASTROCYTES

DIMITRA GEORGOPOULOU

This dissertation is submitted to University College London for the examination of Doctor of Philosophy

University College London, 2015

DECLARATION

I, Dimitra Georgopoulou, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I have indicated this in the thesis.

Dimitra Georgopoulou

ADDITIONAL CONTRIBUTIONS

The work, which was included in the thesis and has been performed by others, is the bioinformatic analysis of Hi-C and ChIP-seq datasets done by Dr. Wen-Ching Chan and Dr. Christopher Barrington and the Hi-C library preparations in neural stem cells done by Dr. Sevil Sofueva.

ACKNOWLEDGEMENTS

I would like to start by expressing my gratitude to Cancer Research-UK for investing in me and thereby allowing me to find my place at the UCL Cancer Institute for the past five years.

I would also like to thank my supervisor, Dr. Suzana Hadjur, for the constant support and meaningful advice. She has been a mentor to me and to the whole 'genome organisation and function' team, guiding us through scientific discovery and to our personal growth with the same enthusiasm as the first day I met her.

The help and support I received from the past and current members of the 'genome organisation and function' group, Sevil, Matteo, Jose, Christopher and Wen-Ching was unmatched. I would especially like to thank Sevil for teaching me the 'C' method and being my good friend. I would like to thank Matteo, my 'PhD-twin', for sharing with me this wonderful course of science and friendship. He has always been there for me sharing knowledge and kindness. Jose and Christopher have always made my PhD days more cheerful with their jokes and more meaningful with their help.

I am also thankful for having all of my friends to support me throughout my PhD. I would like to thank Christina, Athina, Yiannis, Arianna, Berna and Rosina for being my 'rocks' for almost a decade now. I would also like to say thank you to Katerina and Nancy and all my colleagues at the Cancer Institute for sharing a lot of sleepless nights in the lab with me.

Last but not least, I am really thankful for having the love and support of my family, my parents, Yiannis and Helen, my brother, Konstantinos, my aunt, Christina, my uncle, Panayiotis and my cousins, Alexandros and Eirini. For as long as I can remember, they have been actively supporting me, my dreams and hopes. They have always been there to share both my great successes and my biggest failures with equal love and kindness.

Without this help and support, this thesis wouldn't be possible. So I want to say a big thank you to all the people who have contributed to its completion.

To my family

ABSTRACT

The cohesin complex has a conserved and fundamental role in chromosome biology as it is involved in sister chromatid cohesion, DNA repair and gene regulation. Mutations in cohesin and its regulators lead to human developmental disorders and cancer, and it is hypothesised that this is due to cohesin's roles in gene regulation. Cohesin mediates gene regulation primarily through the formation or stabilisation of long-range chromatin loops from CTCF, a conserved transcriptional regulator, binding sites. Cohesin-anchored chromatin loops were first described at specific loci to physically tether distal regulatory elements to gene promoters. This thesis investigated whether the roles of cohesin in mediating chromatin loops at individual loci in the genome could apply genome wide, and if so, whether such global cohesin-anchored chromatin loops could contribute to chromosomal organisation. To test this, I established a post-mitotic cell system from mouse so that cohesin proteins could be genetically deleted in a manner that did not interfere with their roles in mitosis. Using this system, cohesin was shown to be an anchor point for chromatin loops of all scales throughout the genome which included both gene loops and loops required for chromosome architecture. Cohesin-deficient post-mitotic cells exhibited global chromosomal decompaction and widespread transcriptional deregulation. While the majority of cohesin chromatin pools were depleted, chromatin domain borders were not abolished and residual cohesin complexes were still bound to chromatin in cohesin-deficient cells. To test whether such residual cohesin complexes may still be mediating long-range chromatin loops and thus maintaining chromosome structures, I used RNAi to knock-down cohesin regulators, such as CTCF and Stag proteins, in cohesin-deficient post-mitotic astrocytes. My results highlight a role for CTCF and Stag proteins in the regulation of cohesin-mediated genome organisation in mouse. Thus, it was shown that cohesin proteins have key roles in genome organisation and functions.

TABLE OF CONTENTS

LIST OF PUBLICATIONS ARISING FROM THIS THESIS

Martynoga, B., Mateo, J.L., Zhou, B., Andersen, J., Achimastou, A., Urbán, N., van den Berg, D., **Georgopoulou, D**., Hadjur, S., Wittbrodt, J., Ettwiller, L., Piper, M., Gronostajski, R.M., Guillemot, F. (2013) Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. *Genes Dev.* 27, 1769-2786.

Sofueva, S., Yaffe, E., Chan, W.C., **Georgopoulou, D**., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organise chromosomal domain architecture. *EMBO J 32*, 3119-3129.

LIST OF ABREVIATIONS

3C	Chromosome Conformation Capture
3D	Three dimensional
4-OHT	4-Hydroxytamoxifen
4C	Chromosome Conformation Capture on Chip
4C-Seq	Circular Chromosome Conformation Capture with Sequencing
5C	Chromosome Conformation Capture Carbon Copy
ABC	ATP-Binding Cassette
APC	Anaphase Promoting Complex
ASTs	Astrocytes
ATP	Adenosine TriPhosphate
BMP4	Bone Morphogenetic Protein 4
C-motif	Signature motif
Cdk1	Cyclin-dependant kinase 1
CdLS	Cornelia de Lange Syndrome
CH	Chromatin-bound
ChIP	Chromatin Immuno-Precipitation
CoATs	Acetyl Transferases
CREs	Cis-Regulatory Elements
Cre-Ad	Cre-Recombinase expressing Adenoviruses
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CT	Chromosome territories
CTCF	CCCTC-binding Factor
DE	Differentially Expressed
E	Enhancer

EGF	Epidermal Growth Factor
ER	Estrogen Receptor
ERT2-Cre	Estrogen receptor fused to Cre recombinase inducible system
ESCs	Embryionic Stem Cells
FBS	Fetal Bovine Serum
FGF	Fibroblast Growth Factor
FISH	Fluorescence *In Situ* Hybridization
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
FSC	Forward Scatter
GFP	Green Fluorescent Protein
GST	Glutathione-S-Transferase
H3	Histone 3
HP1	Heterochromatin Protein 1
HS	Hyper Sensitive
ICR	Imprinting Control Region
iFRAP	inverse Fluorescence Recovery After Photo-bleaching
I	Insulator
IGF2	Insulin-like growth factor 2
IS	Insoluble
KSHV	Kaposis Sarcoma Herpes Virus
LADs	Lamina-Associated Domains
LCR	Locus Control Region
MACS	Model-based Analysis for ChIP-Seq
MCM	Mini-Chromosome Maitenance
MEFs	Mouse Embryonic Fibroblasts
NBD	Nucleotide-Binding Domain
NSCs	Neural Stem Cells

P	Promoter
PANTHER	Protein ANalysis Through Evolutionary Relationships
Pcdh	Protocadherin
Pds5	Precocious dissociation of sisters protein 5
PE	Pair-end
PFU	Plaque-Forming Unit
PI	Propidium Iodide
Plk1	Polo-like kinase 1
PP2A	Protein Phosphatase 2
pRb	Retinoblastoma protein
qPCR	quantitative Polymerase Chain Reaction
RIN	RNA integrity number
SCC	Sister Chromatid Cohesion
Sgo1	Shugoshin1
SHH	Sonic HedgeHog
siNTR	siRNA Non-Targeting Control
SMC	Structural Maitenance of Chromosomes
SE	Salt-Extractable
SOL	Soluble
SSC	Site Scatter
TADs	Topologically Associating Domains
TBE	Tris-Boric Acid-EDTA
TBS	Tris Buffered Saline
TE	Tris-EDTA
TEV	Tobacco Etch Virus
TSS	Transcription Start Site
UCSC	Genome Browser hosted by the Universiy of California

UV	UltraViolet
Wapl	Wings apart-like protein

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1.

INTRODUCTION

CHAPTER 1. INTRODUCTION

1.1. BACKGROUND

The Structural Maintenance of Chromosome (SMC) protein family consists of condensin and cohesin complexes which act as fundamental regulators of chromosome structure and are conserved from bacteria to humans (Wood et al., 2010). Genetic and biochemical studies have demonstrated that SMC proteins have critical roles in DNA replication, chromosome segregation and chromatin condensation during mitosis and meiosis (Wood et al., 2010). Over the last decade, cohesin complexes have been shown to have functions in addition to their roles in chromosome segregation, and include gene regulation and DNA repair (Merkenschlager and Odom, 2013; Nasmyth and Haering, 2009).

STRUCTURE OF THE COHESIN COMPLEX

MOLECULAR ARCHITECTURE OF THE COHESIN COMPLEX

The core cohesin complex consists of four subunits; two SMC proteins, Smc1 and Smc3, an α-kleisin protein, Scc1 (also known as Mcd1/Rad21), (herein referred as Rad21), and a stromalin protein, Scc3 (also known as Psc3/SA/Stag) (herein referred as Stag) (Anderson et al., 2002; Haering et al., 2002; Haering et al., 2004; Losada et al., 1998; Losada and Hirano, 2001; Melby et al., 1998). In vertebrate somatic cells, immunoprecipitation studies have shown that the cohesin complex can contain either Stag1 or Stag2, but not both (Losada et al., 2000; Sumara et al., 2000). These observations imply that multiple versions of the complex can exist *in vivo*.

The structure of these subunits has been best characterised in *S. cerevisiae*, however the deep sequence conservation of these proteins and similar biochemical results imply that cohesin complexes have similar structures in other species (Anderson et al., 2002; Haering et al., 2002; Hirano et al., 2001). Each SMC subunit folds on itself so that an SMC intra-molecular (Haering et al., 2002) anti-parallel (Melby et al., 1998) coiled coil monomer is formed. Haering *et al.* removed the hinge domains of Smc3 and Smc1 to show that the heterotypic interactions between hinges are both necessary and sufficient for the interaction between the two SMC proteins. More electron microscopy

and biochemical experiments on the bacterial, yeast and vertebrate cohesin show that, at one end, each SMC monomer has a globular hinge domain where the SMC proteins heterodimerise and due to the angle of dimerisation, the two arms create a V-shaped molecule (Anderson et al., 2002; Haering et al., 2002; Hirano et al., 2001). On the other end, each SMC homodimer creates a functional adenosine triphosphate (ATP) nucleotide-binding domain (NBD), when the N-terminus of the Walker A motif (i.e. the phosphate-binding loop motif) dimerises with the C-terminus of the Walker B motif (Haering et al., 2002; Haering et al., 2004; Löwe et al., 2001) (**Figure 1a, b,** Interface 'A'). When ATP is bound to Walker A and B motifs within the Smc1 head, it also binds to a ATP-Binding Cassette (ABC) signature motif (i.e. LSGGQ or C-motif) within the Smc3 head and vice versa, creating a transient bipartite ring that can bind and hydrolyse ATP, a function required for cohesin chromosome binding (Arumugam et al., 2003; Weitzer et al., 2003) (**Figure 1a, b,** Interface 'A').

The cohesin ring closes when the Smc3 coiled coil region near the NBD and the Smc1 NBD bind to the N- and C-terminal domains of Rad21, respectively (Arumugam et al., 2003; Gruber et al., 2003; Haering et al., 2002; Gligoris et al., 2014; Huis in 't Veld et al., 2014) (**Figure 1a**). Haering *et al.* used purified yeast Smc3/1 and Rad21 to show that they co-immunoprecipitate and that the removal of both heads of Smc3/1 but not the replacement of the Smc hinge abolished Rad21 binding to the complex. Also a small linker between Smc3 N- and C-terminus co-purified with Rad21 which supports the arguments that cohesin is a ring (Haering et al., 2002). Another part of the cohesin complex is Stag subunit (Haering et al., 2002; Toth et al., 1999) and it interacts with two regions at the C-terminus of Rad21 (Roig et al., 2014). Crystal structure studies have shown that Stag is a crescent-shaped protein composed of tandem α-helices (Roig et al., 2014) (**Figure 1a**). The C-terminus of Stag, which binds to Rad21, is the region that differs between Stag1 and Stag2 paralogues. The N-terminus is highly conserved and present in organisms lacking other regulators (i.e. Pds5, Wapl, and Scc2/4) (Roig et al., 2014), highlighting its evolutionary conserved roles in cohesin biology. Through mutations or depletion experiments, it was shown to be essential to cohesin loading and release from chromosomes and therefore, to its functions in Sister Chromatid Cohesion (SCC) (discussed below) (Hauf et al., 2005; Hu et al., 2011a; Remeseiro et al., 2012a; Rowland et al., 2009; Toth et al., 1999). This has been further validated when the yeast cohesin complex and the loading of cohesin to chromatin reaction were recapitulated *in vitro*, revealing that Stag is essential for cohesin loading and it stimulates cohesin ATPase activity (Murayama and Uhlmann, 2014).

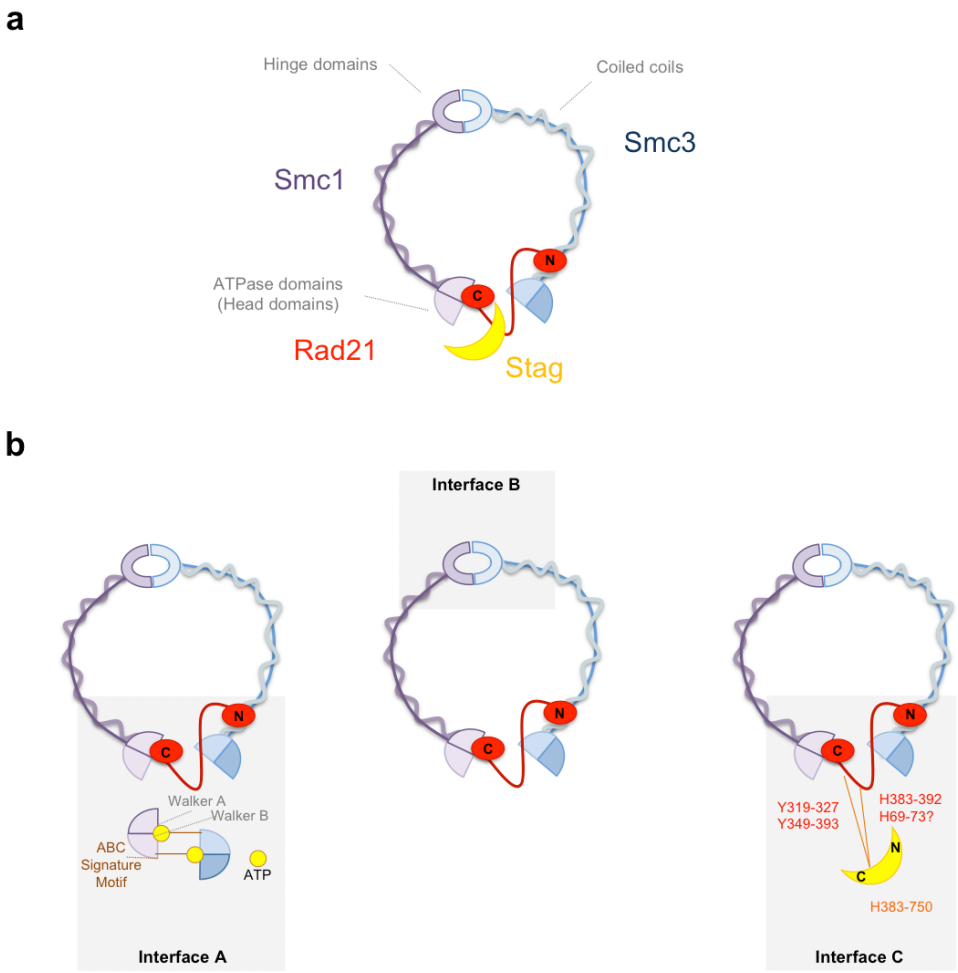

Figure 1. The architecture of the cohesin complex. a) Graphical representation of the core cohesin complex (Smc1 and Smc3, Rad21 and Stag). The hinge region where Smc3 and Smc1 dimerise, the Smc coiled coil region and the ATPase domains (or head domains) are indicated (grey lines). **b)** Interface A depicts the structure of the Smc ATPase domains (i.e Walker A/B and ABC signature motifs) and their association with ATP. Interface B depicts the hinge region where Smc1 and Smc3 dimerise. Interface C presents the amino acid sequence of Rad21 (red), which associates with Stag and vice versa (orange). Adapted from Nasmyth et al. (2011), Arumugam et al. (2003), Zhang et al. (2013) and Roig et al. (2014).

Protein-protein interactions between the core cohesin subunits, specifically the Smc3-Smc1 hinge domains (Haering et al., 2002; Hirano and Hirano, 2002), and SMC ATPase domains-Rad21 interfaces (Arumugam et al., 2003; Haering et al., 2004; Weitzer et al., 2003) are required for cell viability, implying that only intact cohesin complexes are essential for its functions.

Specifically, the disruption of SMC head-Rad21 interface (**Figure 1b**, Interface 'A') leads to decreased viability due to the inability to form cohesin complexes or to

associate stably with chromatin. Initial experiments in yeast showed that mutations at either Rad21 or Smc1 head disrupt their association and result in lower viability (Haering et al., 2004). The chemical fusion of Rad21 to either Smc1 or Smc3 (through its N- or C-terminus) rescues lethality from mutation in the N- or C-terminus of Rad21, supporting the argument that the interactions forming the cohesin ring are essential for cohesin functions (Gruber et al., 2006) (**Figure 1b**). Specifically, mutations at the SMC Walker A/B motifs inhibit cohesin's ability to bind ATP, while mutations at the SMC C-motif prevent ATP hydrolysis and both mutations compromise viability (Arumugam et al., 2003; Weitzer et al., 2003). These studies show that Smc1 but not Smc3 ATP-binding mutants fail to form complexes while both Smc1 and Smc3 ATP-hydrolysis mutants are loaded onto chromatin but accumulate at centromeres. Recent evidence sheds light on why Smc3-Rad21 interaction doesn't require ATP binding. The Rad21 N-terminal region associates with the coiled coil region of Smc3, an interaction fundamental for cohesin association with chromatin and cell viability (Gligoris et al., 2014; Huis in 't Veld et al., 2014). These studies provided evidence for an additional functionally essential interaction within the Smc3-Rad21 interface, implying an additional mechanism that ensures the ring structure of cohesin.

Specific residues at the Smc1/3 hinge region (**Figure 1b**, Interface 'B') are shown to be important for its dimerisation (Haering et al., 2002). This event is essential for the *de novo* association of cohesin with chromatin and cell viability as shown by experiments in yeast using a system where the Smc1/3 hinge regions are linked in a rapamycin-dependent manner (Gruber et al., 2006). *In vitro* DNA binding assays show that SMC mutants that fail to dimerise via the hinge, have decreased ability to associate with DNA (Hirano and Hirano, 2002). Mutations that specifically neutralised a positive charged channel in this interface (i.e. Smc hinge region) led to loss of SCC without affecting Smc dimerisation or cohesin association with chromosomes (Kurze et al., 2011). Interstingly, the above mutations led to a decrease in the acetylation of Smc3, which is an important step during S-phase (discussed below). Thus, it was proposed that this interphase is implicated in conformational changes in the cohesin complex during S-phase and establishment of SCC.

The interface between Rad21-Stag (**Figure 1b**, Interface 'C') is also essential to cohesin functions. Mutations in yeast Stag led to decreased SMC and Rad21 levels on chromatin and vice versa (Toth et al., 1999) and premature sister chromatid separation (Rowland et al., 2009). The yeast Stag is essential for cohesin loading onto chromatin by stimulating cohesin's ATPase activity (Hu et al., 2011b; Murayama and Uhlmann,

2014). It is also important for its release since its phosphorylation during the 'prophase' pathway in vertebrates promotes cohesin removal off chromosomes (Hauf et al., 2005). While the Stag paralogs are 85% similar in sequence, it has been shown both by RNAi-depletion and conditional alleles in mice, that lacking one or both alleles of the Stag genes leads to different functions in cohesion and in DNA repair (Canudas and Smith, 2009; Kong et al., 2014; Remeseiro et al., 2012a). Since cohesin association with DNA is regulated by the cohesin interfaces to which Stag proteins are bound, the different cohesin-Stag complexes may influence differently the way cohesin binds DNA.

COHESIN FORMS A RING-LIKE STRUCTURE WHICH ENTRAPS DNA

The protein-protein interactions described above assemble the core subunits into a tripartite ring-like structure with a diameter of ~ 40 nm (Haering et al., 2002), a structure also visualised by microscopy studies (Anderson et al., 2002). Due to the nature of the cohesin structure and the fact that release of cohesin from yeast chromosomes requires opening of the cohesin ring via cleavage (Uhlmann et al., 1999), it was hypothesised that the complex may entrap DNA inside its ring (Haering et al., 2004). Direct evidence to support this idea came from experiments in yeast, where cohesin subunits remain associated with each other, but not with circular mini-chromosomes, after Tobacco Etch Virus (TEV)-mediated cleavage within the coiled-coil domain of Smc3 or in Rad21 (Gruber et al., 2003). Subsequently, it was shown that when cohesin was bound to circular minichromosomes in yeast, the way to disrupt this association was either by opening the cohesin ring or by restriction enzyme-mediated digestion of the circular DNA (Ivanov and Nasmyth, 2005). Finally, when all of cohesin interfaces were chemically fused or cross-linked together and a protein denaturation step was added to the method, cohesin remained associated with circular, but not linear DNA (Haering et al., 2008). All of these studies supported the argument that cohesin forms a tripartite ring which topologically embraces DNA. These observations led to the question of how cohesin actually associates with DNA. Several subsequent studies explored where the cohesin ring physically opens to accommodate DNA 'entry' and 'exit' and whether one or more cohesin rings are needed to mediate SCC.

THE COHESIN 'ENTRY' AND 'EXIT' GATES

At one end, the cohesin ring is closed by the interactions between the Smc1 and Smc3 hinge regions (**Figure 1b**, Interface 'B'). The nature of the hinge interactions prompted the idea that this interface may function as the DNA 'entry gate' (Haering et al., 2002). Further studies in yeast showed that ATP binding brings Smc1/3 NBD together and therefore, it was suggested that its subsequent hydrolysis enables their topological separation from the hinge (Weitzer et al., 2003). This view was corroborated from studies where fusion of SMC NBDs with N- and C- terminus of Rad21 permitted SCC but prevented cohesion establishment and not its maintenance (Gruber et al., 2006). Mutations in the cohesin hinge domains that don't interfere with SMC dimerisation compromise the acetylation of Smc3 (Kurze et al., 2011), supporting a role for the SMC hinge in cohesion establishment and therefore, an entry of the DNA from this interface.

From the other end, Rad21 functionally brings together Smc1 and Smc3 NBDs (**Figure 1b**, Interface 'A') (Haering et al., 2004), which are shown to act as the 'exit gate' of DNA molecules from the cohesin ring. In mammalian cells, it has been shown that cohesin is removed from chromosomes in a two-step process during mitosis (Losada et al., 1998; Sumara et al., 2002; Waizenegger et al., 2000). During prophase, cohesin removal takes place by the opening of the ring through Rad21-Smc3 interphase and is mediated by Wapl (discussed below). During anaphase, both in yeast and mammalian cells, cohesin removal requires Rad21 proteolytic cleavage at two sites from a cysteine protease, namely separase, in both human and yeast cells (Hauf et al., 2001; Uhlmann et al., 1999). The expression of non-cleavable forms of Rad21 leads to anaphase-arrested cells. Mutations that prevent the phosphorylation of Rad21 during mitosis, enhance the cleavability of Rad21 by separase (Hauf et al., 2005), implying this post-translational modification on Rad21 acts as a protective mechanism from separase-cleavage. In studies investigating cohesin association with chromatin, these cleavage sites were targeted and used as a useful experimental tool that results in the opening of the cohesin rings (Gruber et al., 2003).

COHESIN ASSOCIATES WITH DNA AS ONE OR TWO RINGS

One of the proposed models of cohesin association with chromatin, namely the 'embrace model', predicts that both sister chromatids are entrapped inside the same cohesin ring. Initial immunoprecipitation experiments in yeast showed that the cohesin

complex contains 1:1:1 ratio of the core components (Haering et al., 2002). Ivanov *et al.* developed a method where complexes of circular 'cohesed' mini-chromosomes and cohesin were visualised to show that cohesin was released from circular 'cohesed' mini-chromosomes by either DNA digestion or cohesin cleavage via the SMC coiled coils or Rad21 (Ivanov and Nasmyth, 2005). Then, Haering *et al*. introduced cysteine pairs that could be chemically fused at different cohesin interfaces. When they fused all cohesin interfaces, cohesin remained associated with circular 'cohesed' mini-chromosomes even after a denaturation step (Haering et al., 2008). Later, by using the same methodology, it was established *in vivo* that cohesin is topologically embracing yeast 'cohesed' cohesin (i.e. Smc3 acetylated) (Gligoris et al., 2014).

Other models, the so-called 'handcuff model' propose that a single chromatid is entrapped by a single cohesin ring, and that the two cohesin rings then interact with one another to connect the two sister chromatids (Huang et al. 2005; Zhang et al. 2008b). By biochemical experiments in mammalian cells, chromatin-bound Flag-tagged Rad21 (or Smc1/3) could be immunoprecipitated with Stag 6xHis-tagged Rad21, implying that there could be interactions between cohesin complexes. However, Stag subunits couldn't immunoprecipitate with themselves and their depletion disrupted Rad21-Rad21 interaction. Therefore, it was suggested that two cohesin rings interact with each other via Stag subunit. Later, it was shown that the disruption of the interaction between Rad21 and Stag1 didn't interfere with Stag-Pds5 interactions (Zhang et al., 2013). Since Stag proteins interact with only one region in Rad21, it was proposed that the cohesin 'handcuffs' are stabilised by factors such as Pds5. A study in yeast shows that the depletion of Stag or Pds5 compromised cohesion without interfering with cohesin levels on chromatin (Kulemzina et al., 2012). Thus, Pds5 and Stag depletion is sufficient to destabilise the handcuff while not causing the unloading of the rings (Kulemzina et al., 2012). However, there is not sufficient *in vivo* evidence to conclude whether the topological entrapment of sister chromatids uses embrace, handcuff or a combination of the two models.

THE ROLE OF COHESIN IN SISTER CHROMATID COHESION

SISTER CHROMATID COHESION

Three decades ago, two models for the mechanism by which sister chromatids were 'cohesed' were proposed. One suggested that a physical intertwining of DNA after replication could lead to cohesion of sister chromatids and the second introduced the idea that there are proteins involved in creating SCC (Koshland and Hartwell, 1987; Murray and Szostak, 1985). Koshland *et al.* showed that in yeast chromomosomes DNA catenation is unlikely the main player. By using cell cycle mutants in budding yeast, the authors showed that the majority of cells released from anaphase didn't show detectable DNA catenation on the yeast minichromosomes and the chromosomes segregated properly after mitosis (Koshland and Hartwell, 1987). Initial evidence for a protein involved in SCC during mitosis came from genetic screens in yeast designed to identify factors which failed to maintain SCC after mitotic arrest, identifying several cohesin proteins, such as Rad21 and Stag yeast homologues, Scc2, Eco1, Smc1 and Smc3 (Guacci et al., 1997; Michaelis et al., 1997; Toth et al., 1999). It was also shown that these proteins, which some of them comprise the cohesin complex, function to tether sister chromatids together from DNA replication during S-phase until the end of mitosis when sister chromatids segregate into daughter nuclei (Gruber et al., 2003; Guacci et al., 1997; Losada et al., 1998; Michaelis et al., 1997; Uhlmann et al., 1999). This function for cohesin in SCC is evolutionarily conserved, and found to be essential in yeast, flies and vertebrates (Guacci et al., 1997; Losada et al., 1998; Losada et al., 2000; Michaelis et al., 1997; Vass et al., 2003). Thus, cohesin has an essential role during the cell cycle in maintaining the structure of sister chromatids and forming the bridges that holds them together up to the moment that they are due to be separated to the daughter nuclei.

COHESIN LOADING ON DNA

The loading of cohesin onto chromatin takes place after mitosis and depends on ATP hydrolysis and the cohesin loading complex (**Figure 2**) (Hauf et al. 2005; Nishiyama et al. 2010; Losada et al. 2014; Peters et al. 2009). Although the act of cohesin loading onto chromatin and the players involved are conserved throughout evolution, the timing of cohesin loading differs from yeast to mammals (Ciosk et al., 2000; Hu et al., 2011; Murayama and Uhlmann, 2014, Gillespie and Hirano, 2004; Watrin et al., 2006). In

budding yeast, cohesin is loaded onto the DNA at the end of G_1 phase (Guacci et al., 1997; Michaelis et al., 1997), whereas in mammalian cells this process is initiated already in telophase following reformation of the nuclear envelope (Gerlich et al., 2006; Losada et al., 1998; Sumara et al., 2000). This difference is proposed to be due to differences in mitotic removal pathways of cohesin in these organisms. In budding yeast, most Rad21 is cleaved by separase in metaphase (Uhlmann et al., 1999). In vertebrates, most of the cohesin is removed from chromosomes during the prophase pathway without Rad21 cleavage (Waizenegger et al., 2000), implying that a soluble pool of cohesin is available for immediate re-loading onto the DNA at the end of mitosis.

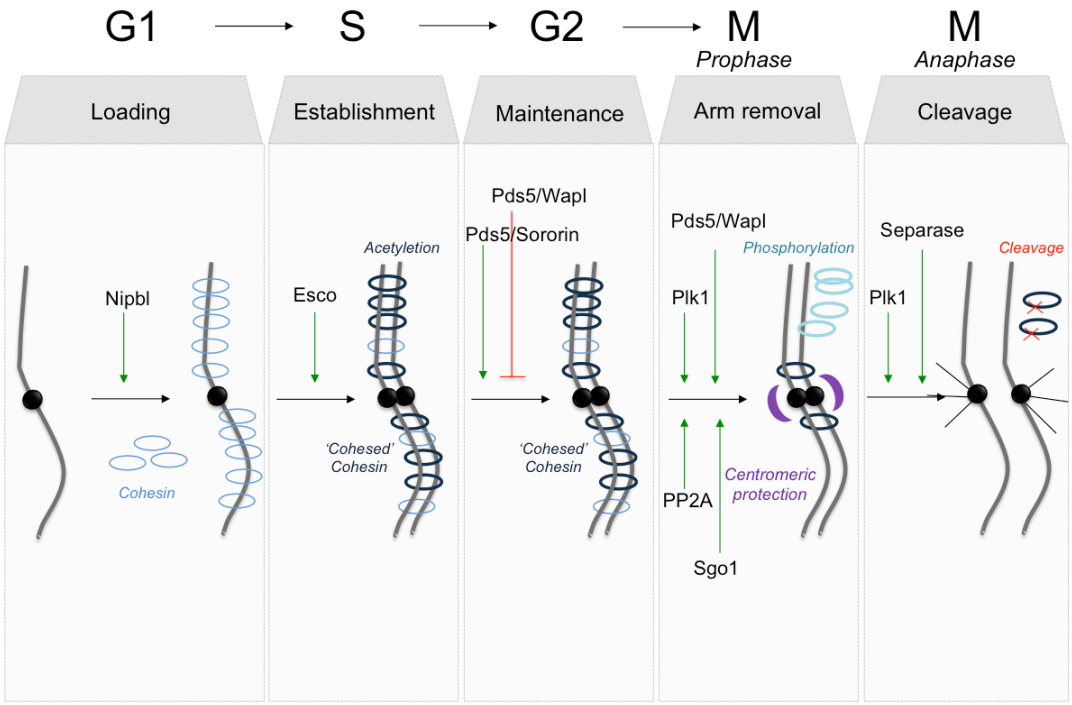

Figure 2. The regulation of the cohesin complex during cell cycle. In G_1-phase, cohesin is loaded on chromosomes, a reaction mediated by the loading complex. In S-phase, Esco proteins acetylate Smc3 and sister chromatid cohesion is established. In mammalian cells, the maintenance of cohesion is regulated by sororin, which is recruited by Smc3 acetylation on Pds5, and by Wapl which antagonises sororin and mediates cohesin removal off chromatin. In mammalian cells cohesin is removed off chromosomes in a two-step process during mitosis. In prophase, Wapl mediates cohesin removal from chromosome arms and the phosphorylation of Stag proteins by Polo-like kinase 1 (Plk1) promotes it. Centromeric cohesion is protected by Shugosin and protein phosphatase 2 (PP2A). At the onset of anaphase, cohesin Rad21 subunit is cleaved by the protease separase so that sister chromatids can segregate to the daughter nuclei. Adapted from Hauf et al. (2005), Nishiyama et al. (2010), Losada et al. (2014) and Peters et al. (2009).

One of the key events required for cohesin loading is the ATPase activity of the SMC NBDs, which was initially shown in yeast (Arumugam et al., 2003; Arumugam et al.,

2006; Ladurner et al., 2014; Weitzer et al., 2003) and suspected to involve the opening of the hinge interface (Gruber et al., 2006) (**Figure 1b,** Interface 'B'). Both Smc3 and Smc1 hydrolysis mutants failed to associate with chromosomes, proposing that ATP hydrolysis is essential for the association of cohesin with chromatin shown in metaphase spreads (Arumugam et al., 2003; Weitzer et al., 2003). *In vivo* studies using ATP hydrolysis mutants showed that they didn't prevent cohesin association with chromatin at core centromeres, where Scc2 (i.e the loading factor) is also enriched, but that these mutants reduced the residence time of cohesin on chromatin (Hu et al., 2011). Corroborative evidence in mammalian cells show that Smc3 ATPase mutants fail to associate stably with chromatin shown by inverse fluorescence recovery after photo-bleaching (iFRAP) (Ladurner et al., 2014). FRAP is used to measure the ability of a molecule to move over time. For this purpose, a fluorophore is covalently attached to the molecule of interest (e.g. protein) and a fluorescence microscope is used to visualise the fluorescently tagged molecule. While standard FRAP follows the recovery of fluorescently tagged proteins into the photobleached region of interest, iFRAP involves the entire fluorescently labelled area being bleached apart from a small region where the loss of fluorescence from this point is monitored over time. I*n vitro* reconstitution experiments in yeast show that cohesin does associate with chromatin transiently but the reaction requires ATP (Murayama and Uhlmann, 2014). Thus, ATP hydrolysis is not required for the ring formation but is fundamental for the stable binding of cohesin on chromosomes.

Several studies underline the conserved functional importance of the Scc2/Nipbl and Scc4/Mau2 complex in the loading of cohesin on chromatin via direct protein-protein interactions in yeast (Ciosk et al., 2000; Hu et al., 2011; Murayama and Uhlmann, 2014), in Xenopus extracts (Gillespie and Hirano, 2004), and in mammals (Watrin et al., 2006). Mutations in Scc2 (Michaelis et al., 1997) or Scc4 (Ciosk et al., 2000) lead to impaired cohesin association with chromatin and defects in SCC, while they don't interfere with cohesin complex assembly. In HeLa cells expressing cohesin ATPase mutants, cohesin levels on chromatin were decreased shown by biochemical and immunofluorescence experiments (Ladurner et al., 2014). This was proposed to be due to their decreased time of residence on chromatin, which was even further reduced after the depletion of the loading factor as shown by iFRAP (Ladurner et al., 2014). I*n vitro* reconstitution experiments in yeast show that cohesin does associate with chromatin in the absence of Scc2/4, however the addition of the loading factor makes the reaction more efficient and stable (Murayama and Uhlmann, 2014). Once loading

has taken place it doesn't interfere with SCC establishment or maintenance (Lengronne et al., 2006). Thus, it is proposed that the cohesin loading reaction involves two steps; a) the initial dynamic association of cohesin with chromatin mediated by Scc2/4 loading factor and then, b) ATP hydrolysis of the SMC NBDs, producing a more stably bound cohesin. The molecular details of how exactly the loading reaction happens remains poorly understood. Some of the outstanding questions in this context include, whether the entrapment of DNA by cohesin involves transient hinge dissociation as well as whether and how the Scc2/4 complex participates in this process.

SISTER CHROMATID COHESION 'ESTABLISHMENT' DURING S-PHASE

After cohesin complexes are loaded on chromatin, either by the end of G_1 in yeast or in early G_1 in vertebrates, there is an important event that needs to take place, namely the establishment of SCC, which transforms cohesin into its "cohesed" state. The 'establishment of cohesion' reaction requires chromatin-bound cohesin, the activity of specific acetyl transferases (CoATs) and must occur specifically in S-phase (**Figure 2**).

Early results showed that while the presence of cohesin on DNA is necessary for SCC (Lengronne et al., 2006), it is not sufficient (Skibbens et al., 1999; Toth et al., 1999; Uhlmann and Nasmyth, 1998). But what is the timing for SCC establishment during cell cycle? SCC can only be established during S-phase, since in yeast when components of the cohesin complex are expressed in G_2 or M phase, they are able to associate with chromosomes but they fail to establish SCC (Haering et al., 2004; Lengronne et al., 2006; Strom et al., 2004).

SCC establishment results in more stable cohesin complexes on chromatin. Evidence for this comes from FRAP experiments performed in yeast (Kurze et al., 2011), in fly (Gause et al., 2010) and in vertebrate (Gerlich et al., 2006; Kueng et al., 2006) cells, that have shown two distinct pools of cohesin after S-phase; one pool with a short and another pool with a long-half-life of cohesin on chromatin. In mammalian cells, Stag2, one of the core cohesin subunits, was tagged with a green fluorescent protein (GFP) and its residence time on chromatin was measured by FRAP (Gerlich et al., 2006). Gerlich *et al.* used Stag2 instead of another cohesin component to infer the time of residence of the cohesin complex on chromatin because GFP-Stag2 was shown to co-immunoprecipite and co-localise with the endogenous Stag1, Smc1, and Rad21 on

5%–20% sucrose-density gradient. This study revealed that cohesin has a mean residence time of 16-25 minutes in both G_1 and G_2 synchronised cells (30-40% of nuclear cohesin) and only G_2 cells possess an additional cohesin pool (30% and 50% of nuclear and chromatin-bound cohesin, respectively) with residence time of ~6 hours. The latter is proposed to be the 'cohesed' cohesin pool. Recently, these observations were further expanded upon studies in G_1 and G_2 sorted HeLa, where cohesin binding sites remained unchanged (Wendt and Peters, 2009), suggesting that the increase in stability in G_2 might not affect the positions where cohesin is enriched across the genome. Interestingly, an additional cohesin 'transient binding' pool in G_1 with a residence time of one minute was observed. This pool was not dependent on the loading factor (Ladurner et al., 2014), raising the question of whether this new transient pool could be similar to the transient cohesin-DNA association that was reported in vitro in yeast (Murayama and Uhlmann, 2014). Thus, it is suggested that cohesin is loaded onto the DNA in G_1 and a proportion of it is 'cohesed' in S-phase with a much higher time of residence and is maintained through G_2.

SCC establishment also requires the activity of specific acetyltransferases, called Eco1 (also known as Eso1/Esco) (**Figure 2**). This function is conserved from yeast to humans (Hou and Zou, 2005; Skibbens et al., 1999; Tanaka et al., 2000; Toth et al., 1999). It was initially shown that an Eco1 loss-of-function mutation leads to premature separation of sister chromatids and lethality (Skibbens et al., 1999; Toth et al., 1999). The specific substrates of Eco1 activity are a pair of highly conserved lysines (K112 and K113 in yeast; K105 and K106 in humans) within the NBD of Smc3, close to but not at, its ATP binding pocket (Ben-Shahar et al., 2008; Nishiyama et al., 2010; Rowland et al., 2009; Unal et al., 2008; Zhang et al., 2008) (**Figure 2**). In budding yeast, replacement of both residues to non-acetylatable forms, causes lethality due to cohesion defects, while loss-of-function mutations of Eco1 are suppressed after the replacement of one lysine on Smc3 to an acetylation mimic residue and cells escape cell death (Ben-Shahar et al., 2008; Rowland et al., 2009; Unal et al., 2008). These results confirm that the two Smc3 lysines are the relevant targets of Eco1 for SCC establishment. In mammals, there are two enzymes, Esco1 and Esco2, which both led to cohesion defects upon depletion (Hou and Zou, 2005; Nishiyama et al., 2010), but only Esco1 was reported to cause acetylation defects (Zhang et al., 2008), implying additional functions for Esco2. During the cell cycle, the acetylation of Smc3 increases during S-phase and is at least partly dependent on DNA replication (Ben-Shahar et al., 2008), which again strengthens the argument about cohesion being established during

S-phase. Therefore, Eco1-mediated acetylation of Smc3 is essential to cohesion establishment and SCC.

SISTER CHROMATID COHESION 'ANTI-ESTABLISHMENT' AND 'ESTABLISHMENT' FACTORS

The role of Smc3 acetylation was unclear until evidence from studies in yeast and mammalian cells showed it counteracts the 'anti-establishment' activity of other cohesin regulators and thus allows sisters to stay 'cohesed' (**Figure 2**) (Ben-Shahar et al., 2008; Chan et al., 2012; Rowland et al., 2009; Sutani et al., 2009; Unal et al., 2008). These regulators are discussed below.

Precocious dissociation of sisters protein 5 (Pds5) is an important regulator of cohesin association with chromatin and its ability to regulate SCC (**Figure 2**). Its essential roles in cohesion are conserved from yeast and flies to vertebrates (Carretero et al., 2013; Chan et al., 2013; Dorsett et al., 2005; Panizza et al., 2000). Psd5 essential roles in SCC were shown by loss-of-function mutations, which suppress the lethality caused by Eco1/Eso1 in yeast (Rowland et al., 2009; Tanaka et al., 2001), even in fission yeast where the gene is not essential for SCC. This predicted a SCC anti-establishment role for Pds5 in yeast. Vertebrate cells contain two Pds5 paralogs, Pds5A and Pds5B. By Pds5A/B depletion it was shown that they are important for SCC in human cells and Xenopus egg extracts (Losada et al., 2005) and that both contribute to telomere and arm cohesion but only Pds5B is required for centromeric cohesion (Carretero et al., 2013). More recent evidence has suggested a two-part mechanism for Pds5's roles in cohesion establishment. The authors showed that Pds5 promotes Smc3 acetylation via its interaction with Rad21 N-terminus during S-phase and protects Smc3 from de-acetylation during G_2/M phases (Chan et al., 2013). In mammals, Pds5 was also proposed to act as 'docking platform' for other cohesin regulatory proteins, sororin or Wapl (discussed below) that contribute to the maintenance or removal of cohesin on or off chromatin, respectively (Nishiyama et al., 2010).

Cohesin complexes are also regulated by wings apart-like protein (Wapl), which was initially discovered for its roles in heterochromatin organisation in *Drosophila* (Verni et al., 2000) (**Figure 2**). Wapl is involved in cohesin regulation on chromatin during the cell cycle by regulating its removal from chromosomes (Gandhi et al., 2006; Kueng et al., 2006). Indeed, its depletion leads to a delayed prometaphase, increased levels of cohesin and poorly resolved SCC but doesn't prevent the cells from exiting mitosis

(Gandhi et al., 2006). Indeed, Wapl-null cells exit mitosis but they have less cohesin available to load in the next cycle (Tedeschi et al., 2013). It also reverses the phenotype of precocious sister chromatid separation caused from Rad21 and Esco2-mutations in mammalian cells (Gandhi et al., 2006). Similarly, the depletion of the yeast ortholog, Rad61, suppresses lethality from Eco1 depletion (Rowland et al., 2009). These results support its roles in SCC 'anti-maintenance' activity. During interphase, Wapl was also shown to be important for maintaining the dynamic nature of cohesin on chromatin. In mammalian cells during G_1 (Tedeschi et al., 2013) and G_2 (Kueng et al., 2006), its depletion by RNAi or genetically increases the amount of cohesin on chromatin and increases the time of residence of the 'dynamic' cohesin pools shown by immunofluorescence and FRAP experiments, respectively. This function is also corroborated in yeast cells in G_2 (Chan et al., 2012). Wapl mediates cohesin functions by directly interacting with the cohesin complex and other cohesin regulator proteins. Specifically, in yeast, Wapl interacts with Stag and the loading factor (Rowland et al., 2009). In mammals, Wapl interacts with core cohesin components and Pds5A/B, as shown by immunoprecipitation experiments and mass spectrometry analysis, and its binding to cohesin depends on the presence of Rad21 and Stag1/Stag2 (Gandhi et al., 2006; Kueng et al., 2006). The functional significance of the above interactions was shown by a series of mutations on Wapl (Ouyang et al., 2013). Thus, Wapl is involved in cohesin removal off chromatin during the cell cycle. It is proposed to promote cohesin removal by transiently opening the cohesin ring via its Smc3-Rad21 interface (Buheitel and Stemmann, 2013; Chan et al., 2012; Eichinger et al., 2013). In flies, Smc3's NBD was linked by a short polypeptide to the N-terminal domain of kleisin and so, cohesin could associate with chromosomes but couldn't be removed shown by FRAP (Eichinger et al., 2013), similar results were obtained in human cells (Buheitel and Stemmann, 2013). Its Wapl-dependant removal during prophase became possible when the linker was cleaved (Eichinger et al., 2013). Altogether, these studies propose that Wapl is a conserved cohesion 'anti-establishment' player.

Sororin, a protein conserved in *Drosophila* (Nishiyama et al., 2010), was initially identified as a substrate of the Anaphase Promoting Complex (APC)/Cdh1 and shown to be fundamental for SCC (Rankin et al., 2005) (**Figure 2**). Its roles in cohesin functions were shown when depletion of sororin led to decreased stable cohesin binding without reducing cohesin amounts on chromatin in G_2-phase (Schmitz et al., 2007). In HeLa cells, RNAi-mediated depletion, co-immunoprecipitation and

immunofluorescence experiments showed that sororin was recruited to chromatin-bound, acetylated Smc3 cohesin during S-phase (Nishiyama et al., 2010). In mitotic chromosome spreads, simultaneous Wapl and sororin depletion leads to a phenotype, which is similar to Wapl depletion alone, proposing a role for sororin in SCC which depends on Wapl (Nishiyama et al., 2010). The authors in this study propose a mechanism whereby sororin's association with Pds5 would prevent Wapl's association with Pds5. Thus, sororin is an additional player in protecting Smc3 acetylated 'cohesed' cohesin by not allowing Wapl to associate with Pds5 and therefore, protecting SCC.

Several lines of evidence support a role for Pds5-Wapl in counteracting the cohesion establishment factor Eco1. Yeast genetics have shown that Eco1 loss-of-function mutants can be suppressed by mutations in Smc3, Pds5, Stag and depletion of Wapl (Ben-Shahar et al., 2008; Chan et al., 2012; Rowland et al., 2009; Sutani et al., 2009; Unal et al., 2008). In budding yeast, recent studies shed light on the role of Pds5 whereby Pds5 promotes Smc3 acetylation during S-phase and Wpl1 removes cohesin (Chan et al., 2012) while during G_2, Pds5 protects it via preventing the de-acetylation of Smc3 (Chan et al., 2013). In flies and mammals, a new player sororin is recruited by Smc3 acetylation counteracting the 'anti-establishment' activity of Wapl (Nishiyama et al., 2010; Tedeschi et al., 2013). Therefore, there is an interplay between establishment and anti-establishment mechanisms so that both SCC and the dynamic nature of cohesin on chromatin are maintained.

SISTER CHROMATID RESOLUTION DURING MITOSIS

In eukaryotic cells, the dissociation of cohesin complexes takes place during two phases of mitosis (Losada et al., 1998; Sumara et al., 2002; Waizenegger et al., 2000) in contrast to yeast where all cohesin is removed during the transition to anaphase (Uhlmann et al., 1999; Uhlmann et al., 2000). In eukaryotes, the first wave of cohesin is removed from chromatin during the 'prophase pathway' when the majority of cohesin dissociates from chromosome arms, but not from centromeres (**Figure 2**) (Waizenegger et al., 2000). This process is regulated by Aurora B and Polo-like kinase 1 (PLK1) (Losada et al., 2002; Sumara et al., 2002). PLK1-mediated phosphorylation of the C-terminal domain of Stag1/2 (Hauf et al., 2005) and Rad21 are shown to decrease cohesin association with DNA (Sumara et al., 2002). By expressing 'non-phosphorylatable' mutants of Stag2 and Rad21, it was shown that Rad21 phosphorylation is dispensable for cohesin dissociation from chromosomes in early

mitosis but enhances the cleavability of Rad21 by separase (described below). Meanwhile, Stag2 phosphorylation is essential for the prophase pathway and is not required for separase cleavage (Hauf et al., 2005). Furthermore, Wapl has an important role in cohesin removal during the prophase pathway. FRAP analysis of Wapl-depleted cells uncovered a reduction in cohesin disassociation from chromosomes during prophase whereas Wapl overexpression leads to precocious loss of SCC (Gandhi et al., 2006; Kueng et al., 2006; Tedeschi et al., 2013). The Scc2/4 complex also dissociates from chromosomes during prophase, which may also contribute to the removal of arm cohesin by Wapl (Watrin et al., 2006). Altogether these results suggest that 'prophase pathway'-dependent phosphorylation events are promoting the removal of cohesin which requires Wapl.

The prophase pathway does not remove all cohesin from chromosomes, as doing so would compromise their bi-orientation on mitotic spindles. How does a population of cohesin 'escape' removal during the prophase pathway? Centromeric cohesin is protected from the prophase pathway due to a centromere-specific protein called Shugoshin1 (Sgo1). Shugoshins bind the PP2A holoenzyme, which is also localised to centromeres in mitotic cells (Riedel et al., 2006; Tang et al., 2006). Depletion of Sgo1 leads to cohesin resolution as early as pro-metaphase without activating the separase pathway (**Figure 2**) (Kitajima et al., 2006; McGuinness et al., 2005; Salic et al., 2004). Thus, Sgo1 is predicted to be the 'protector' of centromeric cohesion during anaphase. The mechanism by which Sgo1-PP2A protect centromeric cohesin was unclear until recently. Liu *et al.* showed that the phosphorylation of Sgo1 recruits Sgo1-PP2A to chromatin-bound cohesin while it promotes the de-phosphorylation of sororin (Liu et al., 2013). Thus, the authors propose that the de-phosphorylation of sororin protects centromeric cohesin from Wapl removal of cohesin.

An additional event, which takes place during mitosis, is the de-acetylation of Smc3. In yeast, this is mediated by the histone de-acetyltransferase Hos1 during anaphase (Beckouet et al., 2010) and in mammalian cells by HDAC8 (Deardorff et al., 2012). In cells lacking HDAC8, Smc3 acetylated cohesin was accumulated at the end of mitosis and in the next cohesin loading cycle, cohesin chromatin occupancy was decreased due to accumulation of acetylated cohesin (Deardorff et al., 2012). The authors found that cleaved fragments of Rad21 associated with Smc1/3 in G_1 in HDAC8-depleted cells and proposed that the loading reaction is inefficient with 'cohesed' cohesin complexes. In this study, the phenotype resembled the one caused by Nipbl depletion.

Thus, de-acetylation is a critical part of the cohesin cycle on chromatin and its deactivation leads to cohesin that can't be loaded *de novo* in G_1.

The majority of soluble cohesin is removed during the 'prophase pathway' and therefore is not cleaved by separase at the metaphase-to-anaphase transition (Sun et al., 2009; Waizenegger et al., 2000). This process does not exist in yeast. It is thus proposed that the purpose of the prophase removal pathway is to generate a large fraction of intact (non-cleaved) cohesin complexes which can then be loaded onto chromosomes during telophase rapidly. It has been suggested that this readily available intact cohesin pool is important for other cohesin functions such as in regulating transcription. A clue to support this is that HDAC8 depletion leads to a non-dissolution and accumulation of acetylated complexes which can't be functionally loaded onto chromosomes and causes de-regulation of transcription similar to that caused by Nipbl depletion (Deardorff et al., 2012).

The second 'wave' of cohesin removal from chromatin takes place shortly before the onset of anaphase, when residual cohesin dissociates due to cleavage of Rad21 by separase (**Figure 2**) (Hauf et al., 2005; Kumada et al., 2006; Nakajima et al., 2007; Uhlmann et al., 1999; Uhlmann et al., 2000). Once chromosomes are bi-oriented on the mitotic spindle, the spindle checkpoint is inactivated and APC/C becomes active. This leads to the ubiquitination and degradation of its substrates, which include the separase inhibitor, securin and cyclin B1, the activating subunit of Cyclin-Dependant Kinase 1 (Cdk1) and thus, activating separase (Waizenegger et al., 2000). Next, as soon as separase is active, it was shown that separase cleaves Rad21 at two sites, which is both required and sufficient for sister chromatid separation (Uhlmann et al., 1999; Uhlmann et al., 2000). The cleavage of centromeric cohesin by separase is a very important event to ensure proper chromosome segregation since either the ectopic expression of non-cleavable Rad21 mutants or the de-activation of separase cleavage led to defects in chromosome segregation (Hauf et al., 2001; Kumada et al., 2006).

A ROLE FOR COHESIN IN GENE REGULATION AND GENOME ORGANISATION

GENE REGULATION IN CIS AND METHODS TO UNDERSTAND GENE REGULATION AND GENOME ORGANISATION

Gene regulation has critical roles during development, evolution and disease (Bulger and Groudine, 2011; Kang et al., 2011). Many genes are regulated in a very precise spatio-temporal manner by cis-regulatory elements (CREs), namely enhancers, and insulators, which can be located at a considerable distance away from their target genes (Segal et al., 2008; Spitz et al., 2003). Enhancers can increase transcriptional output from their target-promoters and are bound by trans-acting proteins. On the other hand, insulators oppose either the positive effects of enhancers. A classic example of gene regulation *in cis* is the developmental gene Sonic Hedgehog (*Shh*), whose expression is modulated throughout development by a set of enhancers located up to 1 Mb away in linear distance (Lettice et al., 2002). Another very well characterised example of *in cis* gene regulation comes from the β-globin locus where DNAse I hypersensitive sites (HS) were found to be located tens of kilobases away from the five β-globin genes and were shown to be part of a Locus Control Region (LCR). LCRs differ from the classic definition of enhancers in that they activate expression of genes in transgene assays independent of where they integrate in the genome (i.e position-independent) and in a tissue-specific manner (i.e. in erythroid cells) (Fraser et al., 1993; Grosveld et al., 1987). So it was hypothesised that in erythroid cells, the HS sites control the expression of the β-globin genes by making the chromatin accessible to trans-acting factors responsible for bridging the genes with their cis-regulatory elements. At the same time, the mechanism by which the 'bridging' of these elements (i.e. enhancers and promoters) takes places remained unclear. That was when the idea of 'DNA looping' became popular predicting that the target gene promoter and the regulatory element are in close physical proximity, thus creating an environment appropriate for gene regulation (Griffith et al., 1986).

A direct relationship between gene regulation and spatial organisation in the nucleus was supported by microscopy studies (Fraser and Bickmore, 2007). One microscopy method widely used is Fluorescence *in situ* hybridisation (FISH), which utilises DNA or RNA fluorescently labeled probes to detect specific DNA or RNA sequences in the nucleus (Bickmore, 1999). Using combinations of fluorescently labeled probes, FISH methods allowed for the visualisation of entire chromosomes (i.e chromosome paints) to individual genes within an individual cell. Such microscopy studies showed the existence of chromosome territories (CT) with preferential positions in the nucleus

(Bolzer et al., 2005), whereby gene-dense chromosomes are localised more to the interior of the nucleus and gene poor at the periphery of the CT (Croft et al., 1999). This evidence suggests that chromatin organisation within the nucleus is not random. Studies looking deeper into sub-chromosome structures showed the re-localisation of genes outside their CT in domains of constitutively high gene expression (Mahy et al., 2002) or the close proximity of co-regulated genes (Brown et al., 2006) to suggest a relationship between transcriptional output and proximity. In this context, co-regulated genes such as the β-globin genes, were found to be close in space frequently by FISH (e.g. at nuclear speckles) and, their close proximity was correlated with transcriptional status (Brown et al., 2006). Recently, using the 3D-FISH method and high-resolution microscopy, the *HoxD13* was found to frequenty co-localise with a tissue-specific enhancer in cells expressing the gene during limb development (Williamson et al., 2012). These observations suggested that the topology of the chromatin fibre was not random and could be functionally significant. While microscopy provided key insights at the level of individual cells, its resolution is restricted to the number of sites one can investigate and its sensitivity to the light-capturing capability of the particular microscope used in the respective study.

Higher resolution methods were required to gain further insights into chromosomes organisation and how this is linked to chromatin functions (e.g. gene expression). In 2002, the Chromosome Conformation Capture (3C) method (**Figure 3**) was developed to study genome organisation in higher resolution and output than that by microscopy methods. The premise behind 3C is that when chromatin fragments are 'proximal' in the nuclear space, they will be more likely to be ligated together and could represent interactions between chromatin fragments (Dekker et al., 2002). The integration of 3C with next-generation sequencing (ie. Hi-C, 4C) has made possible the view of chromatin topology at a resolution and throughput difficult to obtain by microscopy. It is worth noting that the chromatin interaction datasets are based on probabilistic models and thus, provide the 'likely conformation' of chromatin in the given cell type. Another consideration of 3C methods is that they are population-based assays and so they don't allow us to distinguish between a chromatin loop occurring in very few cells or one which occurs in the majority of the cells within the population. Taken together with the dynamic nature of chromatin, this is one of the disadvantages of this approach that one should consider when interpreting 3C data.

Importantly, 3C methods allow the detection of the association/proximity of genomic loci across the genome, such as enhancer-promoter interactions, which are important

for gene expression [reviewed in (Sexton and Cavalli, 2013)]. An example of the above is the identification of the globin promoters interacting with the LCR which correlated with the tissue-specific transcriptional output and thus, creating a 'hub' of active transcription (Tolhuis et al., 2002). Later, these methods proposed that the frequent association of cis-regulatory elements with target gene promoters may be an important requirement of gene regulation and not just a byproduct of gene expression changes. For example, changes in the interactome landscape were observed by using 3C methods before any transcriptional or phenotypic changes occured in a timecourse study during differentiation (Apostolou et al., 2013).

In an effort to understand whether such cis-regulatory chromatin loops exist genome-wide, several derivatives of the 3C method were developed [reviewed in (de Wit and de Laat, 2012; Sofueva and Hadjur, 2012)]. For example, Chromosome Conformation Capture on Chip or Circular (4C) (Schoenfelder et al., 2010) or with sequencing (4C-Seq) method (van de Werken et al., 2012), which detects interactions between one known loci in the genome (i.e. bait) and all possible interactor loci, have been used to generate information about spatial networks of co-regulated genes (i.e. globin genes). Chromosome Conformation Capture Carbon Copy (5C) (Dostie et al., 2006), which detects interactions between several known and several known loci, has been used to underline the physical separation of regulatory units by topologically associated domain borders at the organisation of the X-chromosome inactivation centre (Nora et al., 2012). Finally, two global derivatives of 3C include: a) Hi-C which uses a biotin tag to enrich for all chromatin fragments, which were thought to be proximal in nuclear space, and high-throughput sequencing to globally map chromatin contacts (Lieberman-Aiden et al., 2009) and b) ChIA-PET which includes a chromatin immunoprecipitation step for a protein of interest and only detects a subsets of genomic interactions between regions bound by the protein of interest (Fullwood et al., 2009b). The 3C methods revolutionised the way genome organisation was understood and provided an additional, high-throughput toolbox of methods to understand how the genome folds.

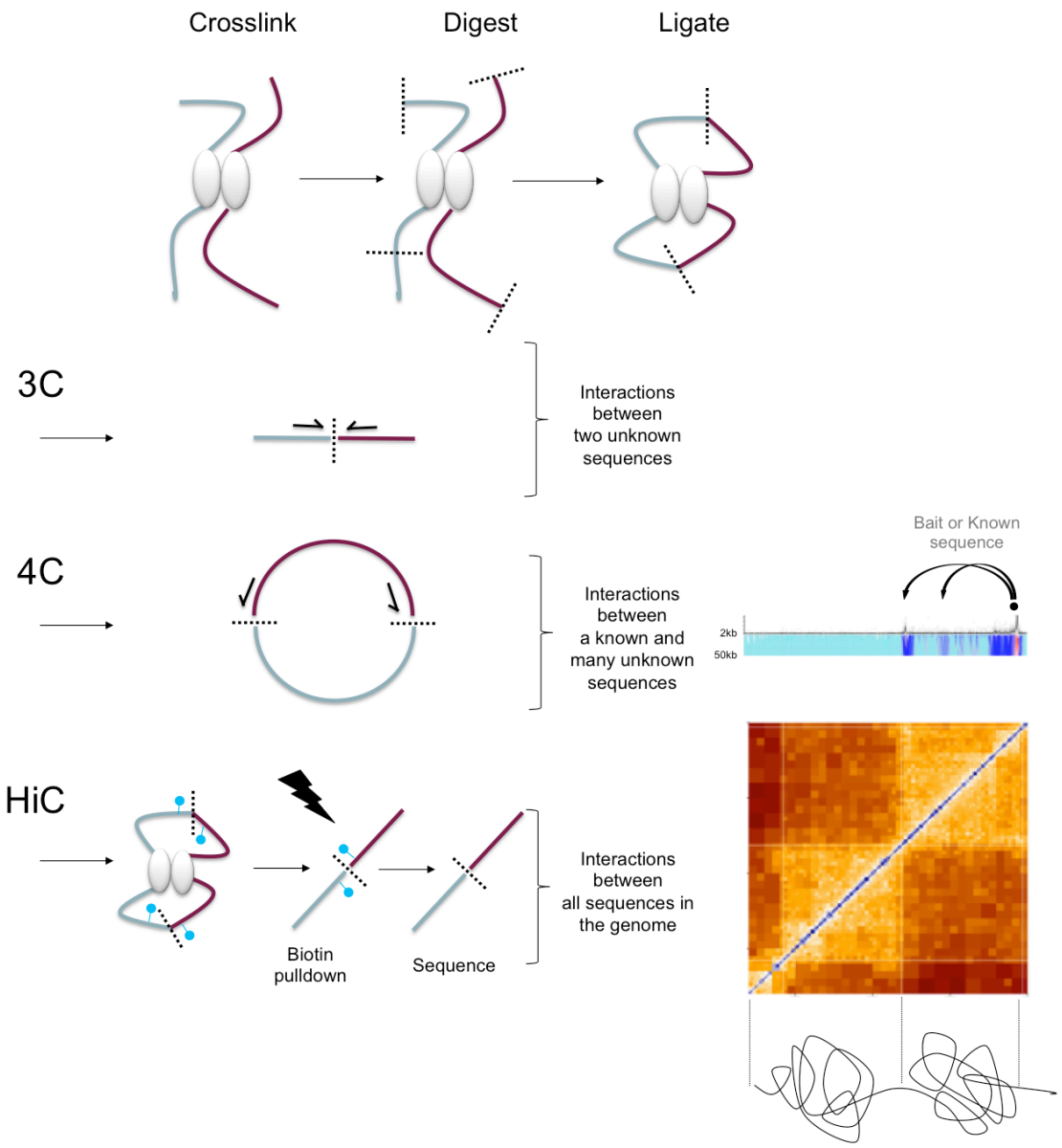

Figure 3. Chromosome conformation capture (3C) method. The principle that the 3C method is based on is that when chromatin fragments are 'proximal' in nuclear space, they will be more likely to be ligated together and could represent interactions between chromatin fragments. 3C techniques include the following main steps: a) crosslinking proteins with DNA using folmaldehyde so that DNA fragments are crosslinked together when they are in close proximity to each other in three-dimensional space within intact nuclei; b) DNA fragments are digested with a restriction endonucleae and ligated so that the closer the fragments are, the more often they will ligate to each other. The 3C approach detects interactions between two known sequences by using primers for both of these sequences. 4C-seq, a candidate approach, uses primers against a known 'bait' sequence to detect interactions with all possible interactor sequences. Hi-C, a genome-wide approach, uses a biotin tag to select all possible chromatin fragments which are found in close proximity in the nuclear space and next generation sequencing to map chromatin interactions genome-wide. Adapted from Sofueva et al. (2012).

COHESIN REGULATES GENE EXPRESSION DURING DEVELOPMENT AND DISEASE

In 1999, a seminal paper in *Drosophila* showed that Nipped-B/delangin, a homolog of the cohesin loading factor, was required for the activation of the Cut and Ultrabithorax homeobox genes, as well as of genes in the Notch pathway (Rollins et al., 1999). The authors were searching for factors that could facilitate the communication between distal enhancers and promoters. Using a screen designed to identify such elements of *cut* expression in *Drosophila*, they showed that the cohesin loading factor, Nipped binds the gypsy insulator element, which is located between the *cut* gene and its enhancer located 85 kb upstream, enabling its activation (Rollins et al., 1999). Later it was shown that cohesin acts as a boundary element separating the *cut* gene and enhancer while Nipped displaces cohesin to activate the *cut* gene (Rollins et al., 2004). Additional evidence to support a role for cohesin in gene expression during development came from studies in zebrafish. It was shown that the reduction in expression of Rad21 or Smc3 impaired Runx gene expression and development of the hematopoietic and nervous systems (Horsfield et al., 2007). These studies provided strong evidence that affecting the dosage of cohesin proteins could result in gene expression defects.

A strong connection between cohesin proteins and developmental gene regulation came from humans. Heterozygous mutations in *NIPBL* and deletions or missense mutations in *SMC1a*, *SMC3*, *RAD21*, and *HDAC8* were reported to be causal for the human developmental disorders Cornelia de Lange (CdLS) and Roberts syndromes (Deardorff et al., 2007; Deardorff et al., 2012; Krantz et al., 2004; Musio et al., 2006; Tonkin et al., 2004). A 30% decrease in Nipbl protein (caused by in-frame mutations) was sufficient to cause severe phenotypes seen in CdLS (Kawauchi et al., 2009). Importantly, cells from patients with CdLS did not exhibit obvious defects in SCC (Vrouwe et al., 2007), raising the question of whether cohesin proteins could have secondary functions in cells. These results also support the notion that cohesin dosage may be important for different functions of cohesin. Indeed, a study in yeast, where the authors achieved titrated cohesin levels, showed that 30% of initial cohesin protein are not sufficient to maintain cohesin function in DNA repair whereas, its reduction to 13% was sufficient to compromise cohesin function in SCC (Heidinger-Pauli et al., 2010). Thus, cohesin dosage is an important factor in an understanding of its different functions.

Evidence for cohesin's gene regulatory role in the pathology of CdLS came when human cell lines were derived from CdLS patients carrying the above mutations exhibited wide-spread transcriptional de-regulation without indications of cohesion defects (Liu et al., 2009), linking cohesin to gene regulation important during human development. Liu *et al.* further supported the direct role of cohesin in gene regulation in the context of CdLS by showing an enrichement of cohesin binding at the transcriptional start sites (TSS) of the de-regulated genes by analysing 15,162 unique transcripts. More recently, in CdLS cell lines carrying loss-of-function mutations in *HDAC8*, cohesin complexes were released from chromatin retaining Smc3 acetylation. These were re-loaded onto chromatin but with decreased occupancy resulting in altered transcription similar to that seen in CdLS cell lines with *NIPBL* mutations (Deardorff et al., 2012). Additionally, mutations in ESCO2 were reported in the CdLS-related syndromes, namely, Roberts syndrome (Schule et al., 2005; Vega et al., 2005). Interestingly, the depletion of other cohesin regulator proteins was also linked to severe developmental defects due to widespread transcriptional de-regulation. Examples include, deletion of Pds5b in mice also results in severe developmental abnormalities in the absence of defects in cohesion (Zhang et al., 2007) and Stag1-null mouse embryonic fibroblasts (MEFs) exhibit transcriptional de-regulation resembling the de-regulation signature of CdLs cell lines (Remeseiro et al., 2012b).

More recently, it has been shown that mutations in cohesin subunits and its regulators are prevalent in several cancers (Leiserson et al., 2015), including glioblastoma (Solomon et al., 2011), myeloproliferative neoplasms (Kon et al., 2013) and bladder cancers (Balbas-Martinez et al., 2013). Supporting the studies in developmental disorders, studies in patients with bladder cancer have shown that *STAG2* mutations are correlated with good prognosis and normal karyotypes (Balbas-Martinez et al., 2013), implying that roles other than those associated with SCC may be defective in this cancer. Similarly, in myeloproliferative neoplasms, *STAG2*-mutated cells were also shown to have normal karyotypes but mutations were correlated with poor prognosis in myeloid neoplasms (Kon et al., 2013). While the impact of cohesin mutations on cancer prognosis and progression are not yet well understood, its roles in gene regulation during development (i.e. *Runx*) (Horsfield et al., 2007), in pluripotency (i.e. *Nanog*) (Nitzsche et al., 2011), and at known oncogenes (i.e. *MYC*) (Rhodes et al., 2010), point to a gene-regulatory function of cohesin in cancer.

COHESIN REGULATES GENE EXPRESSION THROUGH CTCF

Although strong evidence existed for cohesin's role in gene regulation as discussed in the previous paragraph, the mechanism(s) by which cohesin could perform this novel non-mitotic function was unclear. A key clue came with the advent of chromatin-immunoprecipitation (ChIP)-technology, allowing researchers to identify in high resolution, the exact sites on chromatin to which cohesin was bound. The integration of ChIP with genome-wide scale gene expression analyses, led to a deeper insights into cohesin's roles in gene expression.

ChIP studies provided the initial evidence that cohesin may regulate gene expression. ChIP of Scc1 and Smc3 subunits in the yeast genome showed their enrichment both at core centromeres, where it co-localised with the loading factor Scc2, and at distinct sites of active convergent gene transcription along chromosome arms (Lengronne et al., 2004). The authors also showed that the induction of transcription leads to the relocalisation of the cohesin subunits at sites of convergent transcription. Based on these results, it was proposed that RNA polymerase 'pushes' cohesin to relocate from its loading sites to the ones between actively transcribed genes. In *Drosophila* cell lines, ChIP-chip experiments revealed that the binding sites of Smc1, SA and the cohesin loader were detected at actively transcribed genes with RNA polymerase II (Misulovin et al., 2008). In mammalian cells, ChIP-chip and ChIP-Seq experiments also revealed that cohesin was enriched at active transcription start sites across the genome (Parelho et al., 2008; Wendt et al., 2008). Thus, evidence from *Drosophila* and mammalian cells reveals an enrichment of cohesin at actively transcribed genes. Taken together with the fact that cohesin was expressed in post-mitotic cells and tissues (Wendt et al., 2008) strengthened the argument that it could have additional roles apart from SCC.

A key molecular clue about the mechanism by which cohesin regulates gene expression came when cohesin was shown to co-localise on chromatin with CCCTC-binding factor (CTCF), an evolutionarily conserved protein with established roles in transcriptional regulation (Parelho et al., 2008; Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008). ChIP studies of CTCF and Scc1 in mammalian cells showed that the binding profiles for these two proteins were very similar, with ~80% of CTCF sites overlapping with Rad21 and ~60% of Rad21 sites overlapping with CTCF (Parelho et al., 2008; Rubio et al., 2008; Wendt et al., 2008). Co-occupied sites were enriched for transcriptional start sites and sites within gene bodies. Then, it was shown

that cohesin recruitment depends on CTCF. This was revealed by base-substitutions that eliminate CTCF-binding on a known insulator element of the human myotonic dystrophy gene integrated in mouse cells (Rubio et al., 2008), as well as by CTCF protein depletion and ChIP-qPCR studies (Hadjur et al., 2009; Parelho et al., 2008b; Wendt et al., 2008). These observations were confirmed globally using ChIP-Seq (Schmidt et al., 2010) and in other species. In *Drosophila*, ChIP analysis of cohesin binding sites found cohesin co-localised with insulator proteins including with dCTCF, the fly homologue, and the depletion of the latter led to loss of hundreds of cohesin sites (Van Bortle et al., 2014). Biochemical evidence for a direct protein interaction between CTCF and cohesin was provided using Glutathione-S-Transferase (GST)-fusion proteins where the C-terminal region of CTCF was shown to directly interact with both the Stag1 and Stag2 cohesin subunit (Xiao et al., 2011). CTCF mutants at the C-terminal region failed to recruit Stag proteins on chromatin and compromised CTCF insulator function shown *in vitro* by transgene assays and *in vivo* by the transient expression of the mutants in mouse fibroblasts. These observations have led to a model whereby CTCF is recruited to specific sites across the genome and cohesin is localised to CTCF-bound sites through its direct interaction with CTCF.

CTCF, THE VERTEBRATE INSULATOR, AND GENE REGULATION

What underlies the sites to which CTCF is recruited and what mechanisms modulate its occupancy there? CTCF is an 11 zinc finger DNA-binding protein which recognises a 20 bp asymmetric consensus motif and two additional smaller motifs, upstream and downstream of the core 20bp consensus motif. This was shown *in vitro* at selected sites (Filippova et al., 1996) and by genome-wide ChIP-Seq studies at tens of thousands of sites (Cuddapah et al., 2009; Kim et al., 2007; Nakahashi et al., 2013; Parelho et al., 2008b; Wendt et al., 2008). The CTCF motif is highly conserved between species (Schmidt et al., 2012) and across cell types (Wang et al., 2012). CTCF binds to its consensus motif by the combinatorial use of its 11 zinc fingers (Filippova et al., 1996; Nakahashi et al., 2013). *In vitro* studies using gel shift assays showed that deletions or mutations targeting individual or a group of zinc fingers abolish CTCF occupancy at a number of sites (Filippova et al., 1996). Similarly, mutations targeting different groups of zinc fingers followed by ChIP-Seq show that the core zinc-fingers (i.e. 4–7) are responsible for CTCF binding at ~80% of the core motifs (Nakahashi et al., 2013).

CTCF occupancy is regulated at multiple levels with implications for its functions. The CTCF core motif is flanked by an upstream and a downstream motif that increases CTCF occupancy or decreases it, respectively (Nakahashi et al., 2013; Schmidt et al., 2012), implying that flanking motifs are involved in regulation of CTCF occupancy across the genome. However, CTCF occupancy is regulated at additional levels than that of the actual DNA sequence of its motifs. Since the motif is GC rich, CTCF binding is also regulated by DNA methylation, such that methylation of the motif abolishes CTCF binding at several sites (Bell and Felsenfeld, 2000; Wang et al., 2012) (discussed below). In addition, post-translational modifications of CTCF itself, such as phosphorylation of the C-terminal region, which is at the site where it interacts with the cohesin complex through Stag proteins, can modulate its functions without interfering with its occupancy on DNA (El-Kady and Klenova, 2005; Klenova et al., 2001; Xiao et al., 2011). Thus, DNA sequence and modifications modulate CTCF affinity on DNA and therefore, its functions.

CTCF was originally identified as a repressor of *myc* transcription (Lobanenkov et al., 1990). Several studies then revealed that CTCF can act as an insulator by blocking promoters from regulatory elements *in cis* [reviewed in (Herold et al., 2012)]. A region widely used in transgenic reporter *in vitro* assays, which assess whether an element can block the communication between an enhancer and a promoter of a reporter gene, is derived from the chicken β-globin locus. This sequence carries a DNase I hypersensitive site (5'-HS4), which *in vivo* separates the β-globin locus from neighboring heterochromatin (Chung et al., 1993). First, Bell *et al.* inserted this region between an enhancer and a neomycin-reporter to identify potential proteins (i.e. CTCF) bound to the specific regions of strong insulation (Bell et al., 1999) (**Figure 4**). Since then, CTCF binding sites have been identified across the LCR, which regulates the β-globin genes (Farrell et al., 2002). Another well-characterised region where the insulator function of CTCF is important for gene imprinting, is the Insulin-Like Growth Factor 2 (*IGF2*)/*H19* region (**Figure 5**). The two genes share an enhancer but *H19* is expressed only from the maternal allele and *IGF2* only from the paternal allele. In this, a region called Imprinting Control Region (ICR) of upstream of *H19* and down-stream of *IGF2*, is methylated in the paternal allele and is required for the imprinting of both genes. CTCF was shown to bind across the ICR region and that its binding to this region was responsible for blocking the communication between the *IGF2* promoter and its enhancer in an *in vitro* transgene assay (Bell and Felsenfeld, 2000). The authors also found that the methylation of CTCF binding sites at the ICR region

eliminates the binding of CTCF. This evidence pointed to methylation-dependant CTCF binding and in turn, CTCF mediated regulation of gene imprinting at *IGF2*/*H19* locus. Thus, these initial studies described CTCF's role in insulation with implications in gene imprinting.

At the same time, additional roles of CTCF were shown in gene activation. An example of this is the Protocadherin (Pcdh) cluster of genes, which contains promoter-enhancer contacts essential for the development and function of the nervous system. Each Pcdh cluster consists of a number of constant exons located downstream of its alternative exons and downstream of the enhancer regions. The conditional deletion of CTCF from neurons resulted in the de-regulation of the Pcdh cluster and in functional defects (Hirayama et al., 2012). The depletion of the *Pcdhα* enhancer bound by CTCF resulted in reduced *Pcdhac1* expression in brain and increased expression in tissues that normally exhibit very low levels of *Pcdhac1* (Kehayova et al., 2011). CTCF was also shown to function as a barrier between active and inactive regions, which are marked as such by histones modifications (Cuddapah et al., 2009). Thus, CTCF is an essential regulator of genome functions with diverse roles.

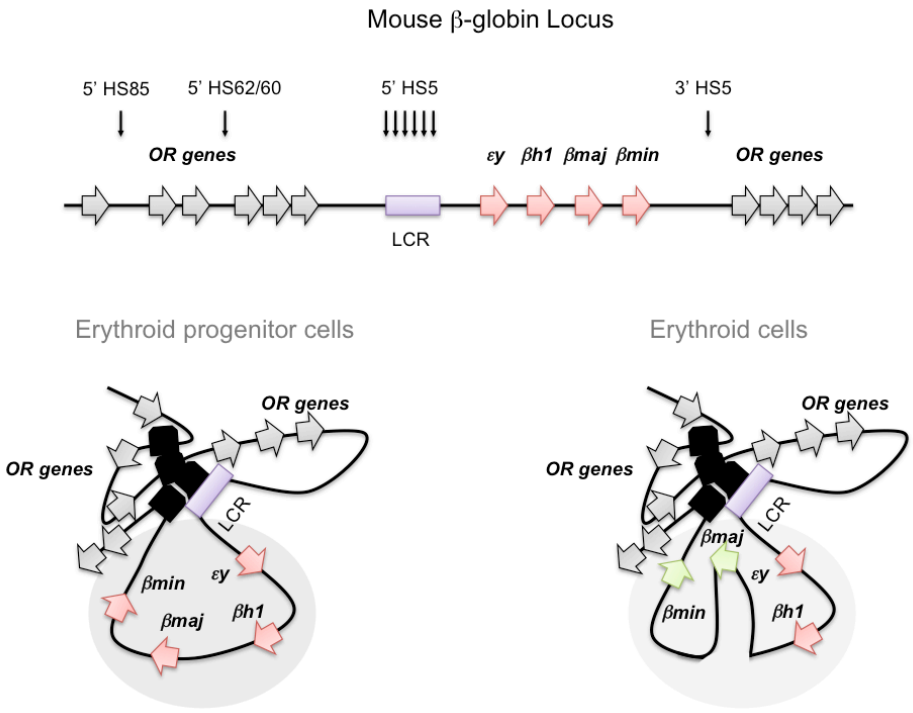

Figure 4. CTCF regulates the conformation of the β-globin locus. Schematic representation of the linear (upper) and 3D conformation (lower) of the mouse β-globin locus. The four b-globin genes (orange arrows) are flanked by the ofractory receptor (*OR*) genes (grey arrows). The expression of the β-globin genes is regulated by a series of *cis*-acting regulatory elements at the 5' and 3' end of the locus (e.g. *εy* and *βh1* in primitive erythroid cells; *β-major* and *β-minor* in definitive erythroid cells). There are six DNase I hypersensitive (HS) sites (black arrows) upstream of the β-globin genes at the Locus Control Region

(LCR) which is required for the high-levels of transcription. There are three CTCF binding sites identified both upstream (i.e. 5′ HS85, 5′ HS62/60, and 5′ HS5) and downstream downstream (i.e. 3′ HS1) of the genes. The lineage-specific CTCF-binding patterns and the 3C-based intrachromosomal interactions are represented as well as the globin gene expression activity (expressed in green and non-expressed in orange arrows) in both erythroid progenitors and erythroid cells. Adapted from Phillips et al. (2009), Tolhuis et al. (2002), Splinter et al. (2006).

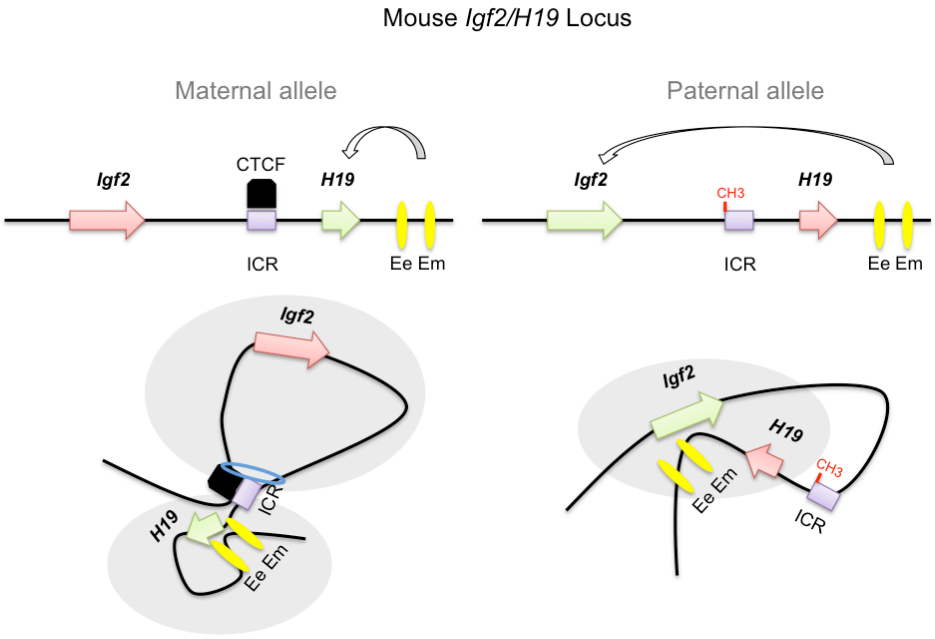

Figure 5. CTCF regulates the conformation and in turn gene expression at the imprinted H19/Igf2 locus. Schematic representation of the linear (upper) and 3D conformation (lower) of the mouse imprinted *Igf2/H19* locus. The maternally expressed *H19* gene is located downstream from the Insulin-like growth-factor 2 (*Igf2*) gene that is expressed exclusively from the paternal allele. The imprinting control region (ICR) upstream of *H19* contains four CTCF-binding sites and is essential for regulation of the locus. Differentially methylated regions (DMRs), such as upstream of *Igf2* promoters and within *Igf2* exon 6, act in concert to regulate reciprocal, allele-specific expression patterns from a set of downstream enhancers, the endodermal tissue enhancer (Ee) and mesodermal tissue enhancer (Em) (yellow ovals) downstream of the *H19* gene. $-CH_3$, DNA methylation. The methylation-dependant CTCF-binding and the 3C-based intrachromosomal interactions are represented as well as the gene expression profiles in both the maternal and paternal alleles. Adapted from Phillips et al. (2009), Kurukuti et al. (2006), Nativio et al. (2009).

COHESIN FACILITATES THE INSULATOR FUNCTIONS OF CTCF

Functional studies show that cohesin contributes to CTCF functions. Wendt *et al.* showed that cohesin is required for CTCF's insulator function at the *IGF2/H19* locus and that this role is independent of its cohesion functions in HeLa cells (Wendt and Peters, 2009). Depletion of either CTCF or cohesin in 293T cells expressing a reporter plasmid carrying the HS4 insulator, revealed a loss of insulation and concomitant

expression of the gene reporter, indicating a role for cohesin in CTCF-mediated insulation (Parelho et al., 2008). Evidence for cohesin and CTCF mediated insulation in viral genomes came from depleting CTCF-binding sites from the Kaposis Sarcoma Herpes Virus (KSHV) genome or cohesin via RNAi in latently infected cells. They showed that cohesin binding is disrupted leading to the non-canonical expression of the lytic cycle gene pointing to a CTCF/cohesin-mediated repression of the lytic cycle (Stedman et al., 2008). Cohesin and CTCF also collaborate to activate transcription. They were shown to mediate alternative promoter usage at the Pcdh family cluster and functional output in neurons (Guo et al., 2012; Monahan et al., 2012). Supporting evidence comes from Stag1-null mouse brains which exhibited de-regulation of the Pcdh cluster suggested to be from loss of Stag1 binding across the cluster shown by ChIP-qPCR (Remeseiro et al., 2012b). Thus, these studies first described cohesin facilitating CTCF insulator and gene regulatory functions.

Cohesin was also shown to regulate gene expression in a CTCF-independent context. While studies looking at co-localisation of cohesin and CTCF by ChIP-chip and ChIP-seq in vertebrates showed that CTCF and cohesin co-localised at the majority of sites across the genome, there was also a number of cohesin sites that were bound to chromatin without CTCF (Rubio et al., 2008; Wendt et al., 2008). New insight as to the functional significance of this group of cohesin sites came from ChIP-seq studies for cohesin and either tissue-specific (Faure et al., 2012; Schmidt et al., 2010) or pluripotency (Kagey et al., 2010; Nitzsche et al., 2011) transcription factors in mammalian cells. Cohesin was shown to be important at these non-CTCF bound sites, since cohesin depletion compromised cell cycle re-entry of breast cancer cell lines as a response to Estrogen Receptor (ER) (Schmidt et al., 2010). The cell cycle re-entry of these cells as a response to ER treatment is thought to arise from transcriptional changes and thus, implying a role for cohesin/ER in transcriptional regulation. Supporting evidence for a functional role for cohesin-no-CTCF sites came from ChIP-Seq studies in primary liver cells, where this group of cohesin sites were co-localising with essential tissue-specifc transcription factors at enhancer regions and thereby, activating tissue-specific transcription (Faure et al., 2012). Another study also shows cohesin-non-CTCF sites at sites of tissue-specific expression in pluripotent ESCs (Nitzsche et al., 2011). These studies suggest a functional role of cohesin in the regulation of tissue-specific gene expression where cohesin-CTCF-independent sites are co-occupied by tissue-specific transcription factors and in turn regulate tissue-specific expression.

COHESIN AND CTCF ANCHOR LONG-RANGE CHROMATIN LOOPS

The fact that cohesin has a conserved role in linking sister chromatids *in trans* during SCC led to the hypothesis that this could apply for cohesin-mediated gene regulation *in cis*. At the same time, cohesin was shown to co-localise across the genome with CTCF, a known chromatin architectural protein. Indeed, CTCF proteins had been shown to be important for mediating chromatin loops *in cis*. The formation of chromatin loops arising from CTCF sites was first shown by (Splinter et al., 2006). Initially, chromatin contacts between the DNase I HS sites at the LCR and promoters of the β-globin were detected in tissues where the genes were active (Tolhuis et al., 2002) (**Figure 4**). Later studies targeting specifically CTCF insulator HS sites at the β-globin locus, via either the conditional deletion of CTCF or targeted disruption of its binding site, destabilised the long-range chromatin interactions that flank the locus and its LCR (Splinter et al., 2006). However there were no changes observed in the transcription of the genes in the locus and a higher-order role in chromatin structure in this region was proposed for CTCF (Splinter et al., 2006). An important study was at the *IGF2/H19* locus where the gene promoters were shown to interact in the paternal allele and chromatin contacts arising from CTCF binding at the *H19* ICR are restricting these interactions in the maternal allele (Kurukuti et al., 2006) (**Figure 5**). So CTCF mediates chromatin loops, which may serve gene regulatory roles, but how CTCF could mediate these chromatin loops was unclear at the time.

Given CTCF's roles in chromatin looping and the colocalisation of CTCF with cohesin all along mammalian chromosome arms, including sites which were known to engage in CTCF-anchored chromatin loops, several groups sought to establish if cohesin could contribute to such CTCF-anchored chromatin loops. The first direct evidence for this hypothesis came from the depletion of cohesin and CTCF and 3C studies at the developmentally regulated gene, *IFNG* in human T-cells. Cohesin and CTCF were enriched at the promoter of *IFNG* as well as distal cis-regulatory elements up- and down-stream of it (Hadjur et al., 2009). Using 3C, chromatin loops could be observed beween the cohesin/CTCF-bound *IFNG* promoter and regulatory elements. Using siRNA, it was shown that these loops were dependent on cohesin but not directly on CTCF, since cohesin depletion didn't affect CTCF binding from these sites. In comparison, siRNA to CTCF led to loss of both cohesin binding and of the contacts. In another example, cohesin and CTCF were shown to mediate contacts at the *IGF2/H19* locus, where their presence is important in the formation of the maternal-specific loops insulating *IGF2* from enhancers downstream and loss of the insulation resulted in

increased expression of *IGF2* (Nativio et al., 2009). In the CTCF-regulated β-globin region, the depletion of the cohesin loading factor Nipbl in mouse and human cells caused reduced chromatin interactions between CTCF- and cohesin-bound insulator regions at the CTCF flanking sites but also within the LCR (Chien et al., 2011). This finding taken with the above examples show that cohesin proteins facilitate CTCF-anchored chromatin loops that may function to regulate gene expression.

In addition to cohesin anchoring chromatin loops from CTCF binding sites, chromatin loops were identified from cohesin-non-CTCF sites. In a CTCF-independent setting, cohesin depletion was linked again to disrupted promoter-enhancer interactions in Embryionic stem cells (ESCs) (Kagey et al., 2010). This study used 3C approaches to also show that subunits of the transcriptional activator complex (i.e. Mediator) and the cohesin loading factor (i.e. Nipbl) were colocalised at cohesin-non-CTCF sites and in turn contribute to cohesin's role in linking tissue-specific enhancers and promoters (Kagey et al., 2010).

Finally, cohesin-anchored chromatin loops have also been reported to structure chromatin for purposes other than gene regulation. Cohesin has a role in mediating chromatin loops important for DNA replication (Guillou et al., 2010). It was observed that cohesin is bound to chromatin at replication origins alongside MCM proteins, factors which are important for replication origin 'licensing'. By using a method they developed, the so-called 'DNA halo' technique, the authors measured the length of the loops at the replication forks. After depletion of Rad21 or Smc3, the length of the chromatin loops were bigger and the inter-fork distances increased. They also reported that S-phase length was increased but without affecting the number of replication forks. This study showed a role for cohesin in mediating chromatin loops at replication origins and influencing S-phase progression.

CHROMATIN LOOPS AND GLOBAL GENOME ORGANISATION

As discussed above, the introduction of 3C and of its global derivatives gave further insight into how gene-regulatory chromatin loops could contribute to the overall compaction of the chromosome. The Hi-C method is an unbiased analysis of global chromosome organisation and is based on the same 'proximity-ligation' principle used for all '3C' techniques (described previously) (Dekker et al., 2002) (**Figure 3**). Hi-C is different from the other 3C methods in that the ligation junctions are labeled with

biotinylated nucleotides, which allows for the enrichment of chromatin fragments which were thought to be previously proximal in nuclear space (Lieberman-Aiden et al., 2009).

The first Hi-C studies confirmed results originally identified by microscopy methods, such as the existence of chromosome territories and open and closed chromatin compartments (Lieberman-Aiden et al., 2009) (**Figure 6**). Due to the increasing resolution of Hi-C methods, new levels of genome organisation were uncovered, which were not possible to identify using microscopy. Chromatin folds so that it creates modular chromatin domains at multiple scales (sometimes called Topologically Associating Domains (TADs) (**Figure 6**). Early definitions suggested that the average size of a TAD was ~ 1Mb in mammalian genomes (Dixon et al., 2012). TADs can be defined as modular units of chromatin that fold into discrete three-dimensional (3D) structures tending to favor internal, rather than external, chromatin interactions. They are delimited by sharp boundaries, which contain housekeeping genes and insulator sites (Dixon et al., 2012). These structures are found in yeast (Duan et al., 2010; Mizuguchi et al., 2014), flies (Sexton et al., 2012), worms (Crane et al., 2015) and mammals (Dixon et al., 2012; Nora et al., 2012), suggesting that they represent a fundamental feature of genomes.

The identification of chromosomal domains led to the question of whether this level of organisation was functionally important. In *Drosophila* embryos, TADs were aligned with ChIP-Seq datasets of 403 epigenetic marks to ask whether domains could represent epigenetic states of chromatin (Sexton et al., 2012). This study identified four classes of domains: the 'Null' domains covered about half of the genome and not associated with any epigenetic marks and have low transcriptional output; the 'Active' domains represented about 40% of the TADs, characterised from active histone marks (i.e. H3K4me3); Two classes of repressive domains, one with Polycomb and one by Heterochromatin Protein 1 (HP1) and repressive histone marks. There was not a correlation between the category of the TADs or gene density but there was with activity, implying their functional roles in transcriptional activity. These findings also apply to mammalian cells where it was shown that TADs correlate with epigenetic marks and therefore, TAD classes were also identified (Rao et al., 2014). The differences in the number and nature of the domain classes could be explained by the available data on epigenetic marks in each study. Further, early studies suggested that TAD borders remain unchanged during differentiation (Nora et al., 2012) or in a

comparison of human and mouse Hi-C maps (Dixon et al., 2012). These early studies suggest that TADs might reflect the functional partitioning of the genomes.

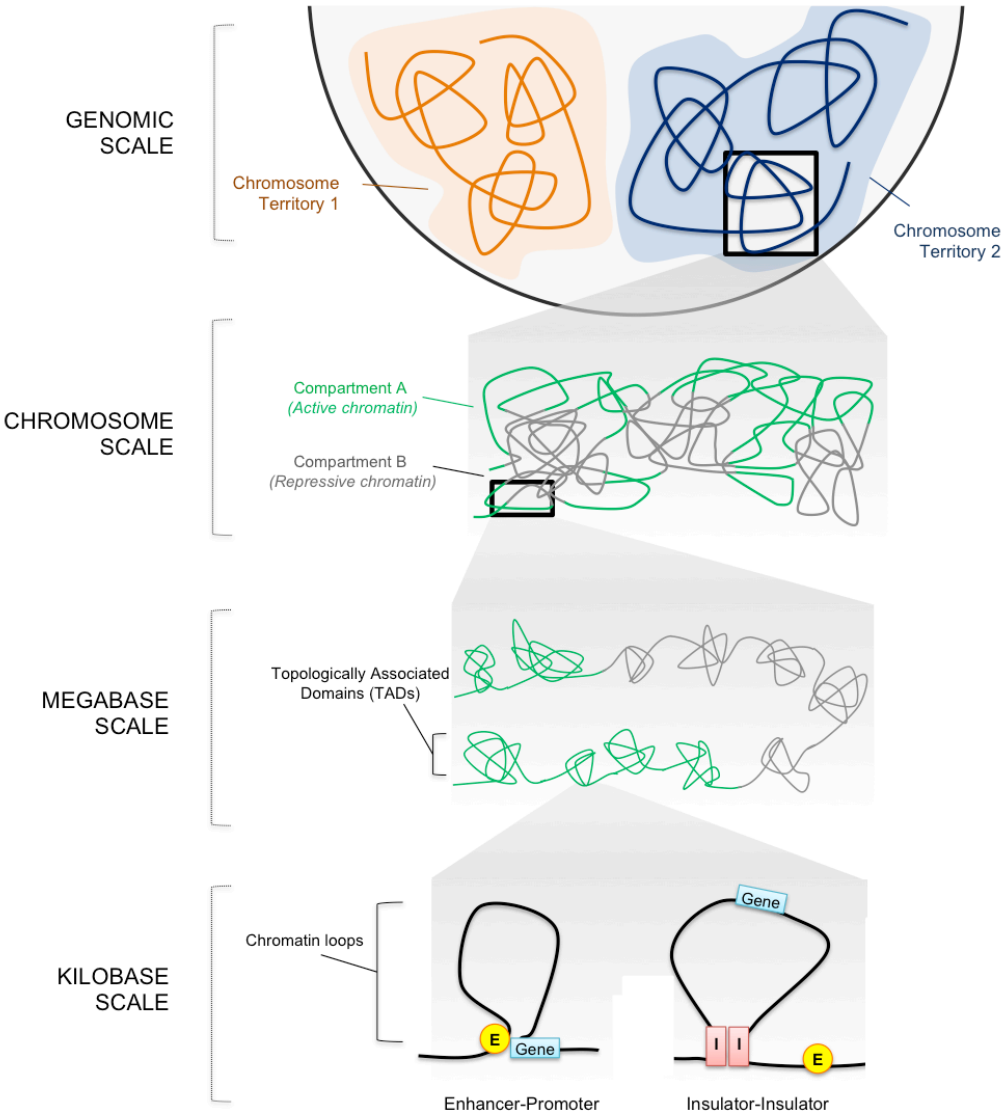

Figure 6. The levels of genome organisation. The chromatin fiber of each chromosome is unraveled to illustrate four different organisation levels [i.e. the chromosome territories, the compartements, the topologically associated domains (TADs) and the chromatin loops]. Chromatin conformations are presented from low (upper) to high (lower) resolutions. The chromatin fiber and the corresponding chromosome territories are shown in bleu and orange. The A and B compartments visualised at a multimegabase scale (Active and repressive chromatin in green and grey, respectively) are shown separately to highlight their inherently distinct nature, although there is no evidence that their conformations differ at the level of TADs (megabase scale). Two examples of chromatin looping (submegabase scale) are shown: 1) enhancer-promoter and 2) enhancer-silencer. E: enhancer; P: promoter; I: nsulator. Adapted from Fraser et al. (2015).

The functional significance of the TAD border region was shown by their depletion. Nora *et al.* used the 5C method at the X-chromosome inactivation center to look for chromatin structure in this region (Nora et al., 2012). They investigated the importance of TAD domain borders by deleting a ~50 kb border region. They showed ectopic contacts between the neighboring domains and transcriptional up-regulation of genes within the domain region by qPCR (e.g. *Jpx*). It was with a recent study by Lupianez *et al.* that the functional importance of borders was more clearly shown. The authors used Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology to generate mouse models for developmental deformities in the limb caused by chromosomal re-arrangements. They did 4C-Seq from promoters of genes located in three different TADs with the one in the middle enclosing an enhancer element. They were able to correlate the chromosomal re-arrangements with changes in the transcriptional regulation *in cis* of genes and consequent phenotypic changes in the developing mouse limb (Lupianez et al., 2015). Importantly, all these changes caused by the chromosomal re-arrangements were accompagned with the disruption of the TADs borders at these regions. For example, a deletion of an extended region crossing a TAD border resulted in ectopic enhancer-promoter interactions, which otherwise were spatially separated into two neighbouring TADs. This ectopic chromatin contacts resulted in the activation of *Pax3*, which is not expressed in the limb and caused developemental defects in the limb (i.e. Brachydactyly-like phenotype). This phenotype occurred only if the depleted region disrupted a TAD border, suggesting the functional importance of borders in spatial gene regulation with implications in development.

If TAD borders are important for maintaining TADs and the maintenance of TADs is important for correct gene regulation, then it is valuable to know which factors are responsible for creating TAD borders. TADs borders are enriched for housekeeping genes, SINEs and the insulator proteins, CTCF as well as cohesin (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). While CTCF was enriched at the TAD borders, most CTCF binding sites were not at TAD borders, but rather within TADs (Dixon et al., 2012), implying that CTCF may not be sufficient for TAD maintenance. However, one should consider that TADs exist at multiple scales in the genome and therefore, CTCF could mediate TAD insulation at several scales and not strictly at the 1 Mb scale (discussed further below). Later, it has also been suggested that the 'strength' of a TAD border correlates not only with the presence of insulator proteins bound, but the number of proteins at a given TAD. In *Drospohila*, insulators have diversified, resulting in a large number of proteins which function in insulation (including Su(hw), Zw5,

CP190, dCTCF, Beaf-32) and found in different combinations across the genome (Schwartz et al., 2012). Using ChIP-seq for a selected number of these factors, it was suggested that those TAD borders, which were most robust, were associated with the largest occupancy of insulator proteins (Van Bortle et al., 2014). Another study raised the possibility that their combinatorial presence across the genome might also be functionally significant (Schwartz et al., 2012). The authors showed that a number of different classes of genomic sites are occupied by specific combinations of insulator proteins proposing that this may confer different functions to this insulator 'spots' across the genome. All these studies implied that CTCF may have a role in TAD structure but also that the number (or combinations) of insulators could also be involved in this.

Increased resolution of chromosome contact maps revealed the complexity of TAD structure. A study using 5C across seven loci encompassing genes important for pluripotency and differentiation (e.g. Oct4, Nanog) and ChIP-Seq for groups of factors (e.g. histone marks, CTCF and cohesin) in ESCs and ESC-derived NSCs gave further insight to TAD structure. The authors proposed that TADs encompass sub-TADs, which were thought to be important for tissue-specific expression since their structure changes (5C contact profiles) between cell types (Phillips-Cremins et al., 2013). The authors also proposed another class of contact profiles at longer-range which was similar to both cell types in this study, namely 'constitutive' loops. The proteins that were found to be enriched at the basis of the 'constitutive' loops were CTCF and cohesin whereas at the 'gene-regulatory' loops, a combination of architectural proteins including Nipbl, the cohesin loading factor was found (Phillips-Cremins et al., 2013).

Additional molecular methods have been developed to look at global genome organisation called CHIA-PET (discussed above). However due to the low coverage nature of CHIA-PET, these methods have not been as successful as Hi-C. Despite this, both cohesin and CTCF ChIA-PET have been performed (DeMare et al., 2013) (Handoko et al., 2011), and broadly these studies have confirmed the results of Hi-C. Such studies either looking from cohesin (DeMare et al., 2013) or CTCF (Handoko et al., 2011) binding sites in the mouse limb buds and pluripotent cells, respectively, captured chromatin loop events shared between cohesin and CTCF. They also proposed from analysing for tissue-specific gene activity and histone mark profiles that cohesin creates loops around actively transcribed genes but also conserved loops between tissues (DeMare et al., 2013) similar to what it was proposed in (Phillips-Cremins et al., 2013).

Altogether the above studies suggested that cohesin and CTCF could have a role in TAD organisation, which are highly conserved structures and functional elements of genome organisation in a plethora of species.

1.2. THESIS AIMS

The cohesin complex is essential for sister chromatid cohesion during mitosis and therefore, fundamental for cell survival. Cohesin has also been shown to function in gene regulation from studies in developmental models. Cohesin was shown to regulate gene expression via mediating chromatin loops at several loci. I hypothesised that cohesin proteins may tether chromatin *in cis*, similar to the example loci, throughout the genome during G_1 and that such cohesin-anchored chromatin loops could have a global role in mediating genome organisation.

To address this hypothesis, I pursued the following aims:

1. To establish a cell culture system where the roles of cohesin in gene regulation could be analysed separately from its essential roles in mitosis. To do this, mouse neural stem cells were differentiated *in vitro* into post-mitotic astrocytes where Rad21, the core cohesin subunit, was genetically depleted by using the inducible Cre-recombinase system (ERT2-Cre) system.

2. To determine if the ability of cohesin to mediate long-range interactions between distal regulatory elements and target gene promoters at select loci existed throughout the genome and how such loops contributed to genome organisation. To do this, a genome-wide cohesin chromatin immunoprecipitation analysis was integrated with chromosome conformation capture (Hi-C and 4C) studies to characterise the cohesin-anchored chromosome structure. Furthermore, to determine the functional role of cohesin in chromosome structure and transcriptome were analysed in Rad21-deficient post-mitotic ASTs.

3. To determine how known cohesin complex regulators contribute to cohesin stability on chromatin and in turn to chromosomal structure in post-mitotic cells. To address this, biochemistry methods were employed to identify the cohesin regulators present in Rad21-deficient ASTs and genome-wide approaches were used to measure the impact of their RNAi-mediated depletion on cohesin stability and chromatin contacts.

CHAPTER 2.

COHESIN-DEFIECIENT POST-MITOTIC ASTROCYTES EXHIBIT GLOBAL DOMAIN DECOMPACTION AND GENOME-WIDE DEREGULATION OF TRANSCRIPTION

CHAPTER 2. COHESIN-DEFIECIENT POST-MITOTIC ASTROCYTES EXHIBIT GLOBAL DOMAIN DECOMPACTION AND GENOME-WIDE DEREGULATION OF TRANSCRIPTION

2.1. INTRODUCTION

Cohesin proteins are essential for sister chromatid cohesion (SCC) from the point of DNA replication until the onset of anaphase (Ciosk et al., 2000; Losada et al., 1998). The complexes are loaded onto chromosomes before S-phase and are removed from chromosomes by the end of mitosis (Ciosk et al., 2000; Rolef Ben-Shahar et al., 2008; Waizenegger et al., 2000). Mutations in the cohesin complex lead to the human developmental disorders Cornelia de Lange (CdLS) and Roberts Syndrome (RS), which are associated with transcriptional de-regulation but not SCC defects (Krantz et al., 2004; Liu et al., 2009; Musio et al., 2006). These results suggested that cohesin proteins could have alternative functions in gene regulation and early studies done in *Drosophila* and zebrafish supported this argument (Horsfield et al., 2007; Rollins et al., 1999).

Despite these observations, it was still unclear how a protein involved in SCC could also influence gene expression. ChIP studies revealed that cohesin proteins colocalise with a known transcriptional regulator, CTCF (Parelho et al., 2008; Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008). One mechanism by which CTCF was known to regulate gene expression was by forming chromatin loops between cis-regulatory elements and target genes (Tolhuis et al., 2002). The widespread colocalisation of CTCF and cohesin plus the known ability of cohesin to tether DNA molecules *in trans* for SCC, made cohesin-anchored chromatin looping an attractive hypothesis for how cohesin could contribute to gene regulation. Indeed, cohesin was shown to regulate gene expression via mediating the formation of chromatin loops that bridge cis-regulatory regions such as promoter-enhancers or enabling the insulator function of CTCF at a number of loci in the genome (Hadjur et al., 2009; Kagey et al., 2010; Nativio et al., 2009). With these observations in mind, I hypothesised that a vast array of cohesin-mediated long-range interactions could contribute to the spatial organisation of the genome. In order to be able to study the function of cohesin in gene regulation, it was necessary to develop a system by which I could eliminate its essential roles in mitosis.

The differentiation of mouse neural stem cells into non-dividing post-mitotic astrocytes (ASTs) *in vitro* represents an optimal system for the investigation of cohesin G_1 functions (Conti et al., 2005). First, neural stem cells can be readily expanded *in vitro* into large cell numbers to accommodate the downstream genomic methods. Second, neural stem cells can be easily converted into post-mitotic cells by the simple addition of a hormone *in vitro* (i.e. BMP4). Third, cohesin subunits can be depleted with genetic tools specifically in post-mitotic cells without the concern that this would influence cell cycle effects. Finally, neural stem cells represent a disease-relevant model which may contribute to the understanding of CdLS, a developmental syndrome presented with mild to severe neurological symptoms (Mannini et al., 2013).

Using this system, I investigated cohesin roles in genome organisation and transcriptional regulation specifically in G_1. I used a multidisciplinary approach incorporating genetic tools, microscopy, transcriptome analysis, ChIP-seq and 3C methods.

2.2. RESULTS

IN VITRO DIFFERENTIATION OF MOUSE NEURAL STEM CELLS INTO POST-MITOTIC ASTROCYTES

A series of clonal neural stem cell (NSC) lines were previously derived from mouse embryonic stem cells (ESCs) carrying a conditional allele of Rad21 gene and the tamoxifen-inducible Cre-recombinase (ERT2-Cre) (referred to as *Rad21*$^{Lox/Lox}$) by Dr S. Hadjur. As a control, NSCs were also prepared which didn't carry the conditional allele for *Rad21* but did have the ERT2-Cre (referred to as *Rad21*$^{WT/WT}$) (**Figure 7**).

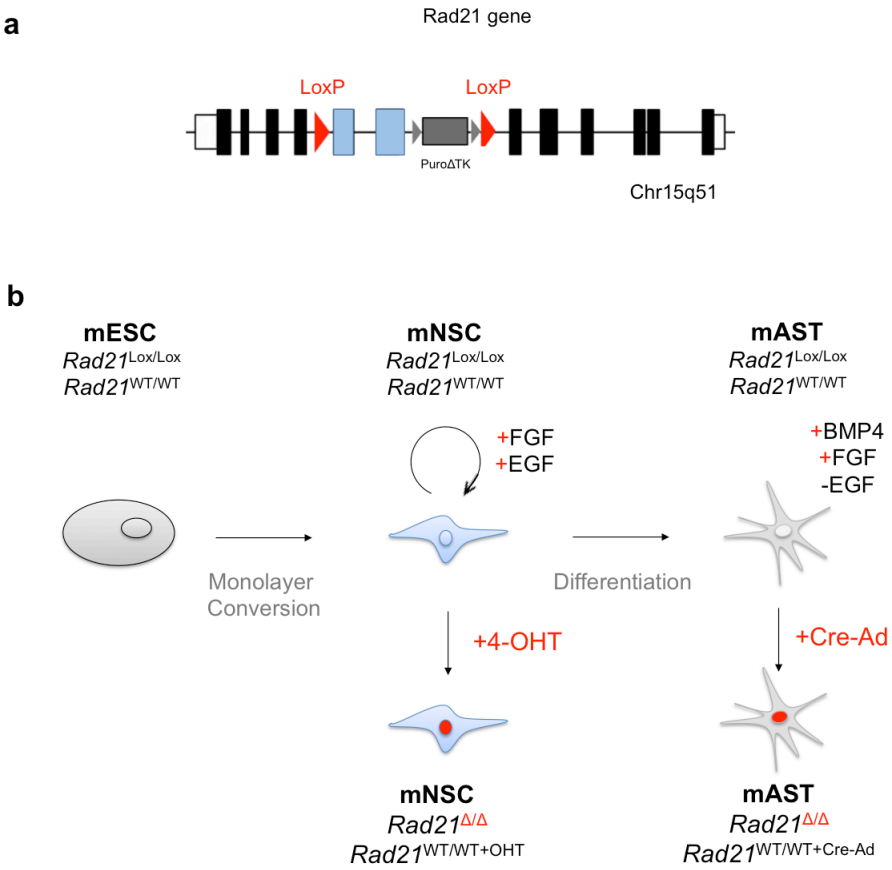

Figure 7. Graphical representation of the mammalian model and the genetic approach. a) Rad21 gene is located on the chromosome 15 and has LoxP sites flanking the 5th and 6th exon (bleu) and a selection cassette for puromycin (grey). **b)** Mouse *Rad21*$^{Lox/Lox}$ and *Rad21*$^{WT/WT}$ embryonic stem cells (ESCs) were converted as monolayers into neural stem cells (NSCs) by EGF and FGF addition and then, NSCs were differentiated into astrocytes (ASTs) by BMP4 addition. *Rad21* depletion was achieved in NSCs and ASTs by 4-hydroxytamoxifen (4-OHT) treatment for 48 hours and Cre-recombinase expressing adenoviruses (Cre-Ad) infection for 96 hours, respectively. Epidermal Growth Factor (EGF), Fibroblast Growth Factor (FGF), Bone Morphogenetic Protein 4 (BMP4).

Thereafter, the above NSC lines were induced to differentiate *in vitro* into post-mitotic astrocytes (ASTs). *Rad21*$^{Lox/Lox}$ and *Rad21*$^{WT/WT}$ NSCs were maintained under EGF plus FGF conditions and were induced to differentiate into ASTs by the addition of

BMP4 plus the removal of EGF and significantly decreased amounts of FGF2 to the culture media (**Figure 7**, differentiation conditions described in CHAPTER 5). The successful conversion of ESCs to NSCs and the subsequent *in vitro* differentiation into post-mitotic ASTs was confirmed by expression of key lineage-specific marker genes (**Figure 8,** 2 clones per cell type). Oct4, Nanog and Sox2 are core transcription factors regulating pluripotency (Silva and Smith, 2008) and were expressed in the mouse ESCs used in this study. NSCs expressed the radial glial markers Blbp and Glast as well as the neurogenic factors Mash1 and Olig2, while ASTs expressed Gfap, Aqp4, Fgfr2, Slc6a8 and Slc1a2 (Conti and Cattaneo, 2010).

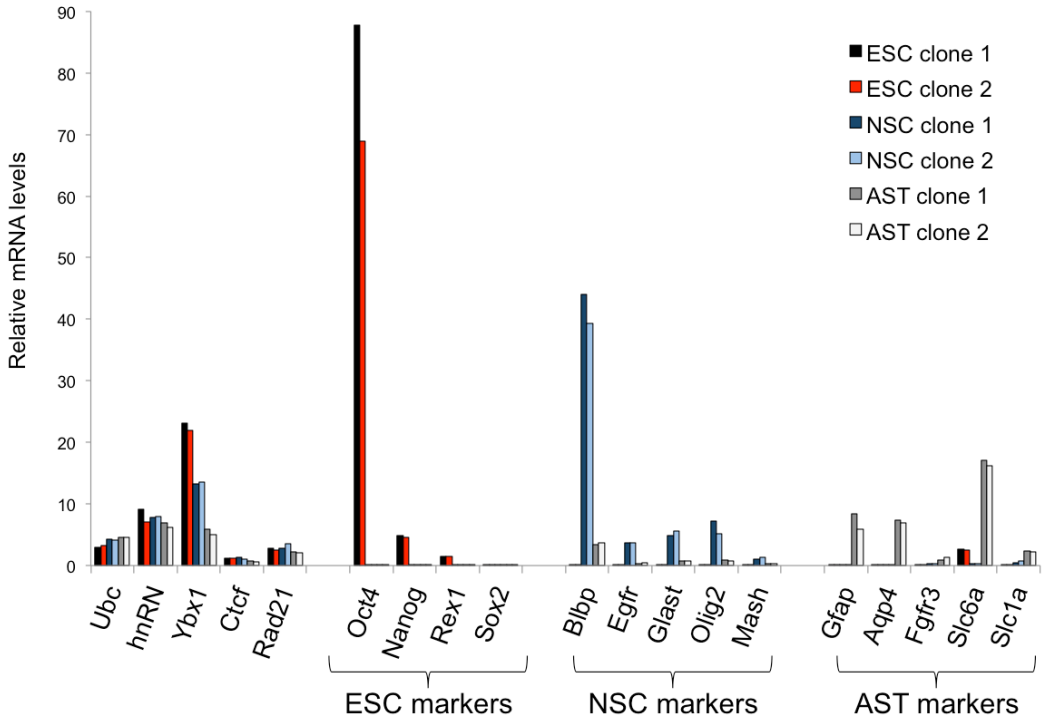

Figure 8. Transcriptional analysis of the *in vitro* ESC conversion to NSC lines and their differentiation into ASTs. Relative mRNA levels by qPCR analysis of housekeeping genes (Ubc, hnRNP K, Ybx1), Ctcf and Rad21 genes as well as lineage-specific genes including ESC-specific (Oct4, Nanog, Rex and Sox2), NSC-specific (Bplb, Glast, Egfr, Mash1 and Olig2), AST-specific genes (Gfap, Aqp4, Fgfr2, Slc6a8 and Slc1a2) in two independent clones of ESC (black and red), NSCs (bleu) and ASTs (grey) *Rad21*$^{Lox/Lox}$ (2 clones per cell type; 1 independent experiment per clone).

In addition to mRNA analysis, the conversion of ESCs to NSCs and their subsequent differentiation to ASTs was confirmed by immunofluorescence experiments to evaluate the expression and localisation of a subset of these key lineage-specific proteins. In agreement with the mRNA analysis, NSCs were positive for Nestin and Olig2 proteins, while ASTs were positive for GFAP (**Figure 9**).

The homogeneity of the ASTs cultures was not evaluated in this study as the aim was to create a non-dividing cell population to isolate the role of cohesin in gene regulation from its roles in cell cycle. The immunofluorescence experiments for markers of cell division, such as Ki67, indicated that Ki67 was expressed in both the ESC and NSC dividing cell populations, but it was absent from ASTs (**Figure 9**). This was further confirmed using Propidium iodide (PI) staining during a timecourse of BMP4 addition to NSCs (**Figure 10a**, 2 independent experiments). This analysis showed that more than 95% of ASTs were enriched in G_1 upon treatment with BMP4 for 24 hours (**Figure 10a**). As cells differentiated into ASTs *in vitro*, several morphological changes were observed, initiating as early as 12 hours post-BMP4 addition (**Figure 10a**).

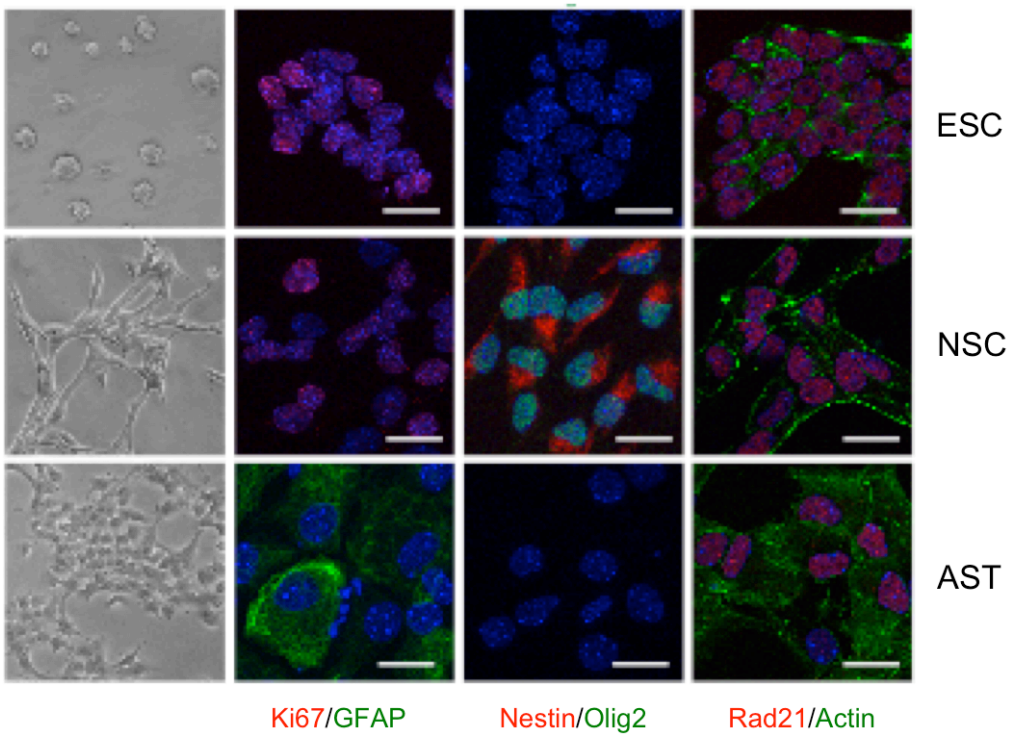

Figure 9. Immunofluorecence analysis of the *in vitro* ESC conversion to NSC lines and their differentiation into ASTs. Bright field images (left panels) of cellular morphology (magnification 20x) and maximum projections of immunocytochemistry and subsequent confocal imaging of ESCs, NSCs and ASTs after 48 hours of BMP4 addition stained for a proliferation marker (Ki67 in red), lineage-specific expression of NSC markers (Nestin and Olig2 in green and red), AST markers (GFAP in green), cohesin subunit and actin (Rad21 and actin in red and green). DNA has been counterstained with DAPI (blue) (Scale bar=20μm).

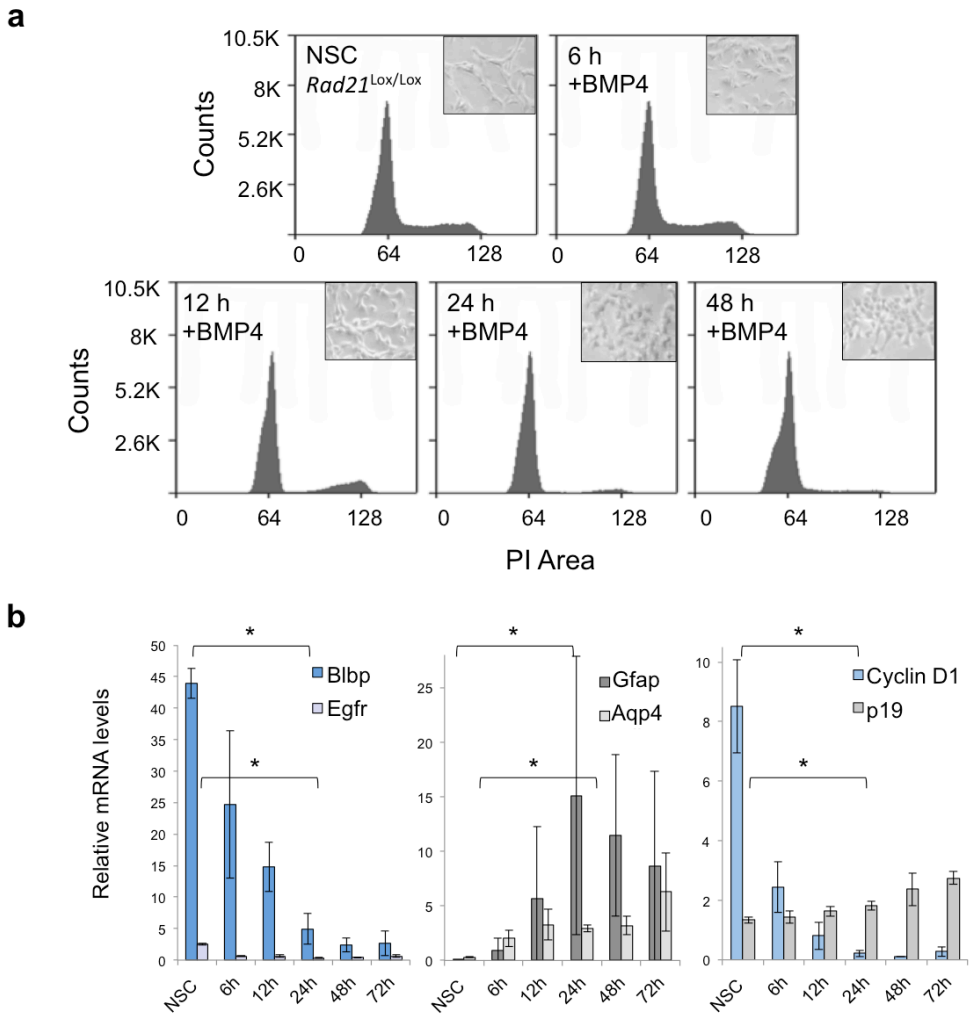

Figure 10. NSCs differentiated into post-mitotic ASTs by BMP4 addition exit the cell cycle and start expressing ASTs-specific genes. a) Distribution of cell cycle phases by flow cytometry and cell cycle analysis in a BMP4 timecourse experiment in *Rad21*$^{Lox/Lox}$ NSCs (10,000 events per sample). Percentages in each cell cycle phase, sub-G1, G1, S and G2/M phases, respectively: NSC (5%, 70%,10%, 15%); +BMP4 6h (1%, 69%, 12%, 18%); +BMP4 12h (2%, 80%, 2%, 18%); +BMP4 24h (2%, 92%, 1%, 5%); +BMP4 48h (8%, 85%, 3%, 4%) (2 independent experiments). Bright field images of cellular morphology (magnification 20x) of the timecourse experiment (top right). **b)** Relative mRNA levels by qPCR analysis including NSC-specific (Blbp, Egfr in blue), AST-specific genes (Gfap, Aqp4 in grey), G1-S transition of the cell cycle promoter and inhibitor genes (CyclinD1 and p19 in bleu and grey, respectively) during BMP4 timecourse in NSCs *Rad21*$^{Lox/Lox}$ normalised to housekeeping genes (*Tbp, Ubc*) (3 independent experiments). Error bars represent the standard deviation and statistical significance was assessed using a Wilcoxon Rank Sum and Signed Rank Test, Blbp and Egfr, Gfap and Aqp4, Cyclin D1 and p19: $p \leq 0.045$ for the indicated timepoints of BMP4 addition.

Future experiments aimed to delete *Rad21* specifically in non-dividing cells. Thus, a differentiation timecourse analysis was performed in order to determine the point after BMP4 addition when NSCs start to express AST-specific markers and it would be appropriate to delete *Rad21*. In addition to the differentiation markers described in **Figure 10b**, genes known to be important during the G_1 to S phase transition of the cell cycle were also assessed (i.e. CyclinD1, a G_1-S transition marker, and p19, a G_1-S

transition inhibitor) (**Figure 10b**, 3 independent experiments, p≤0.045 assesed by Wilcoxon Rank Sum and Signed Rank Test). This data showed that during NSC differentiation by treatment with BMP4, NSC marker genes showed a 8-fold decrease, whereas AST marker genes (i.e. Aqp4) an 8-fold increase in mRNA levels by 24 hours. By 24 hours, cell cycle promoter genes showed a 40-fold decrease in mRNA levels, whereas cell cycle inhibitor genes, which had low basal expression levels, show a 2-fold increase (**Figure 10b,** 3 independent experiments, p≤0.045 assesed by Wilcoxon Rank Sum and Signed Rank Test). This suggested that NSCs had reached a 'post-mitotic-like' state around 24 hours post-BMP4 treatment.

While I could confirm that Rad21 protein levels were detected in dividing ESCs and NSCs as well as in post-mitotic ASTs (**Figure 8, 9)**, in agreement with previous studies (Wendt et al., 2008), I observed a consistent reduction in Rad21 mRNA and protein levels in ASTs after 24 hours of differentiation from NSCs (**Figure 11a,b**, 3 and 2 independent experiments, respectively). This reduction in Rad21 levels remained stable after 96 hours of differentiation (**Figure 11a,b**; 1 experiment in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{WT/WT}$, respectively). The significance of this observation is unknown.

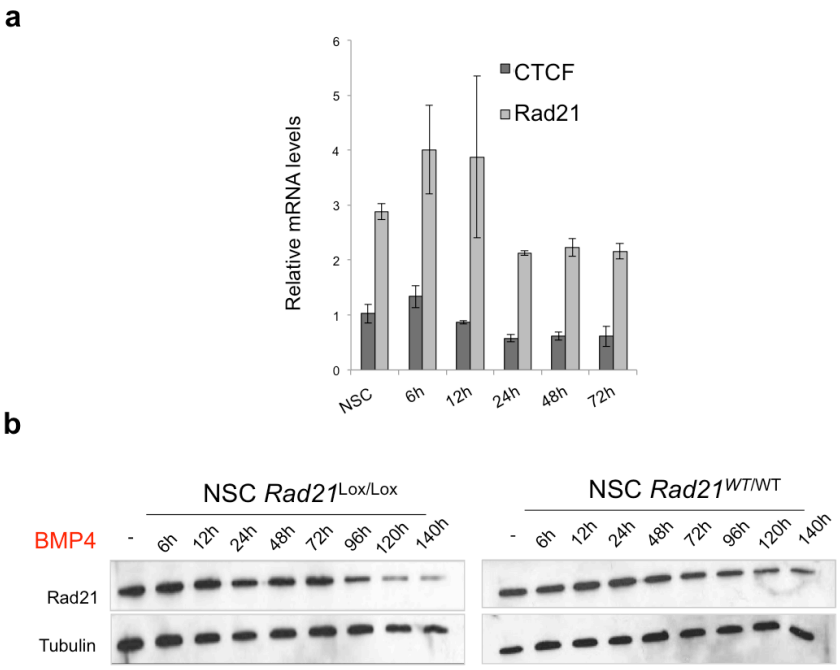

Figure 11. NSCs differentiated into post-mitotic ASTs by BMP4 addition express Rad21, core cohesin subunit. a) Relative mRNA levels by qPCR analysis of Rad21 and Ctcf genes (grey) during BMP timecourse in NSCs *Rad21*$^{Lox/Lox}$ normalised to housekeeping genes (*Tbp, Ubc*) (3 independent experiments). **b)** Western blot analysis for Rad21 and tubulin protein probed on the same membrane during a BMP4 timecourse in both *Rad21*$^{Lox/Lox}$ and *Rad21*$^{WT/WT}$ NSCs (1 experiment per genotype).

In summary, ESCs were converted and cloned to NSC lines expressing markers characteristic of NSC biology that were used to derive *in vitro* AST cultures by adding BMP4. NSC cultures treated with BMP4 for 24 hours already were expressing AST-specific proteins and exited the cell cycle.

GENERATION OF RAD21-DEFICIENT NSCS AND ASTS

To study the functional effects of cohesin on gene regulation, I first optimised the conditions for depletion of *Rad21* in NSCs and ASTs. Loss of Rad21 cohesin subunit was previously shown to be sufficient to remove the complex off chromatin (Hadjur et al., 2009; Parelho et al., 2008). Since *Rad21*$^{Lox/Lox}$ ESCs carried the ERT2-Cre transgene, both NSCs and ASTs were initially treated with various concentrations of 4-hydroxytamoxifen (OHT) to measure the deletion at the level of Rad21 DNA, RNA and protein.

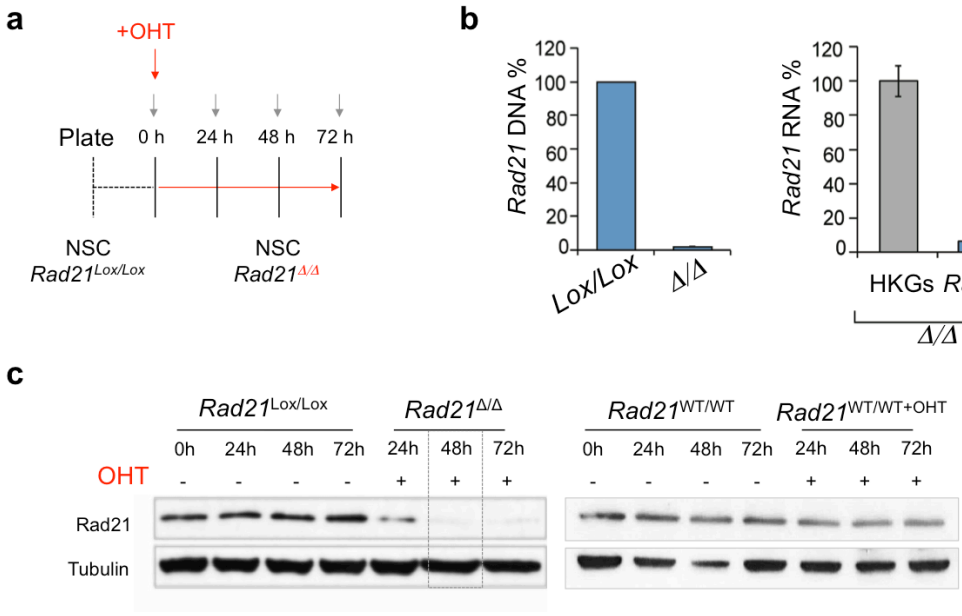

Figure 12. Rad21 is depleted by 4-hydroxytamoxifen treatment in Rad21$^{Lox/Lox}$ NSCs. a) Graphical representation of the protocol used to deplete *Rad21* in *Rad21*$^{Lox/Lox}$ NSCs. NSCs were grown for 24 hours when 4-hydroxytamoxifen (OHT) was added. Grey arrowheads represent timepoints when samples were collected. **b)** Relative Rad21 mRNA and genomic DNA levels by qPCR analysis (3 and 5 independent experiments, respectively) in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ NSCs normalised to housekeeping genes (*Gapdh* and *Ubc, Tbp* respectively). Error bars represent standard deviation. **c)** Western blot analysis for Rad21 and tubulin protein probed on the same membrane during an OHT timecourse in both *Rad21*$^{Lox/Lox}$ and *Rad21*$^{WT/WT}$ NSCs. Grey box indicates the timepoints used in downstream analysis.

NSCs treated with OHT for 48 hours resulted in efficient loss of Rad21 DNA, mRNA and protein (**Figure 12a-c**). Specifically, qPCR primers designed to quantitatively

analyse the extent of genomic deletion showed that more than 95% of genomic DNA was deleted (**Figure 12b,** 3 independent experiments). Similarly, more than 95% of mRNA was reduced upon OHT treatment based on qPCR analysis (**Figure 12b,** 5 independent experiments). Rad21 protein levels were assessed in timecourse experiments of OHT treatment. Rad21 protein levels were considerably lower in *Rad21*$^{Lox/Lox}$ NSCs after 48 hours of OHT treatment whereas Rad21 protein expression remained stable in control *Rad21*$^{WT/WT}$ NSCs treated under the same conditions (**Figure 12c**).

Unexpectedly, ASTs did not respond similarly to NSCs when treated with OHT (**Figure 13a**, 1 experiment). Although Cre-recombinase was detected in the cytoplasm by Western blotting after OHT treatment, its translocation into the nucleus was not sufficient to induce the deletion of Rad21 from the nuclear fraction (data not shown).

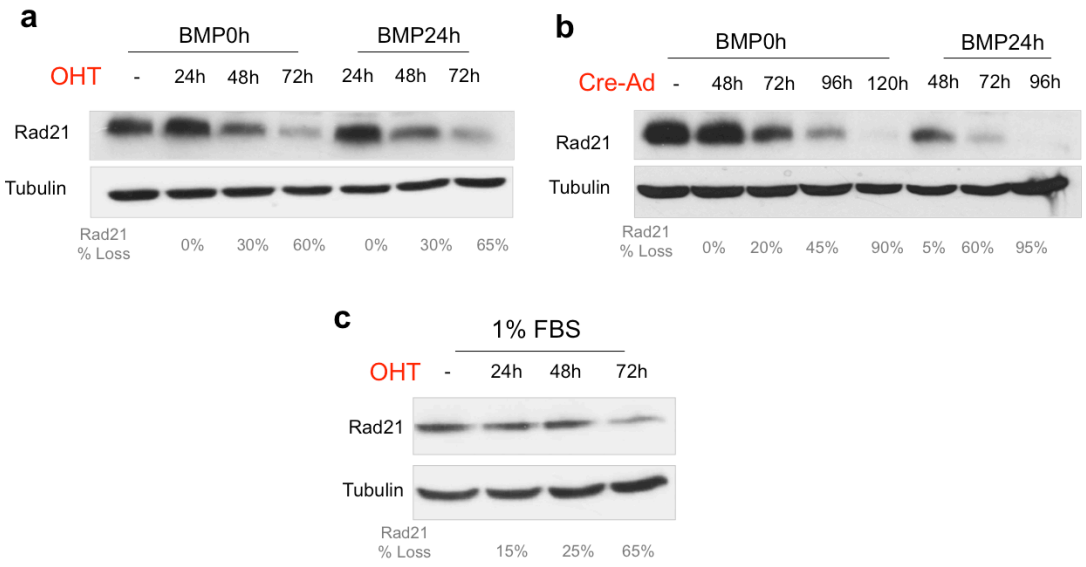

Figure 13. Rad21 depletion is most efficient by Cre-expressing adenovirus in ASTs. a) Western blot analysis for Rad21 and tubulin protein probed on the same membrane during an OHT timecourse in *Rad21*$^{Lox/Lox}$ ASTs where either OHT was added 24 hours post the addition of BMP4 or at the same timepoint as BMP4 (1 expreriment). **b)** Western blot analysis for Rad21 and tubulin protein probed on the same membrane during an Cre-Ad timecourse in *Rad21*$^{Lox/Lox}$ ASTs where either Cre-expressing adenoviruses (Cre-Ad) was added 24 hours post the addition of BMP4 or at the same timepoint as BMP4 (1 experiment). **c)** Western blot analysis for Rad21 and tubulin protein probed on the same membrane during an OHT timecourse in *Rad21*$^{Lox/Lox}$ ASTs where NSCs were differentiated by the addition of 1% Fetal Bovine Serum in the cell culture media (1 experiment). The quantification of Rad21 intensity by ImageJ software, normalised to tubulin levels in each timepoint and percentage of Rad21 loss indicated in grey (values rounded to the nearest 5%).

In order to optimise the conditions under which Rad21 is efficiently depleted from ASTs, the following strategies were tested: a) BMP4-induced differentiation and subsequent OHT treatment which is the control condition when the levels of Rad21 in the controls remained the same (**Figure 13a,** 1 experiment); b) BMP4-induced differentiation and Cre-expressing Adenovirus (here referred as Cre-Ad) addition at the same time as BMP4 addition or 24 hours later (**Figure 13b,** 1 experiment); c) Fetal Bovine Serum (FBS)-induced differentiation and subsequent OHT treatment to explore the possibility that BMP4-induced differentiation affects OHT-mediated depletion of Rad21 (**Figure 13c,** 1 experiment).

Based on these optimisations, the condition chosen for downstream experiments was the expression of Cre-recombinase by adenoviral infection (Cre-Ad) of $Rad21^{Lox/Lox}$ and $Rad21^{\Delta/\Delta}$ AST for 96 hours (**Figure 14**). As before, qPCR was used to assess the extent of genomic DNA deletion and mRNA levels in ASTs treated with Cre-Ad for 96 hours. 95% of Rad21 DNA and RNA was depleted, while 85±5% of Rad21 protein was depleted from whole cell lysates after 96 hours of Cre-Ad treatment (**Figure 14a-c,** 3 independent experiments). Additional immunofluorescence experiments for Rad21 and Actin in $Rad21^{Lox/Lox}$ and $Rad21^{\Delta/\Delta}$ ASTs showed that the majority of cells in the population had lost Rad21 protein. Here, the results present a representative maximum projection image as well as the quantification of Rad21 fluorescence intensity performed by Volocity software. Rad21 intensity is shown as a violin plot to represent the data as a quantile distribution (Nuclei=350 from 3 independent experiments, $p=2.2x10^{-10}$ assessed by Wilcoxon Rank Sum and Signed Rank Test) (**Figure 14d**) and implying that the small amount of protein still detectable on the Western blots is not from cells in the population which have escaped Cre-Ad treatment or received lower amounts of viral particles compared to other cells during the infections. This microscopy analysis showed that ASTs uniformly lose cohesin protein 96 hours post Cre-Ad infections.

In summary, NSCs and ASTs deficient for Rad21 cohesin subunit were generated so that they could be used in downstream experiments to investigate the effects in cohesin deficiency in gene regulation in both dividing and post-mitotic cells.

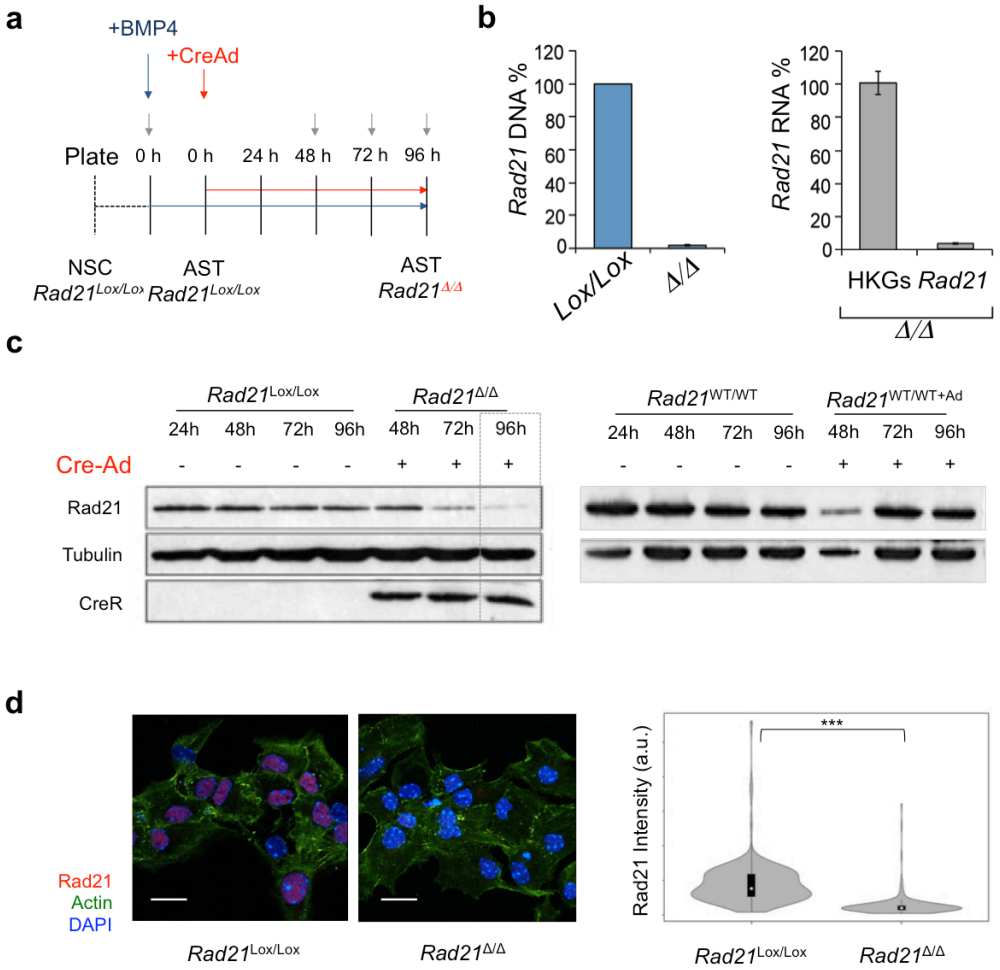

Figure 14. Rad21 is depleted by Cre-expressing adenovirus infections in ASTs. a) Graphical representation of the protocol used to deplete Rad21 in *Rad21*$^{Lox/Lox}$ ASTs. NSCs were plated at optimal confluence and grown for 24 hours when BMP4 was added. Then, ASTs 24 hours were infected with Cre-expressing adenoviruses (Cre-Ad). Grey arrowheads represent timepoints when samples were collected. **b)** Relative Rad21 mRNA and genomic DNA levels by qPCR analysis (3 and 5 independent experiments, respectively) in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ ASTs normalised to housekeeping genes (*Gapdh* and *Ubc, Tbp* respectively). Error bars represent standard deviation. **c)** Western blot analysis for Rad21 and tubulin protein probed on the same membrane during a Cre-Ad timecourse in both *Rad21*$^{Lox/Lox}$ and *Rad21*$^{WT/WT}$ ASTs. At 96 hours post-infections, 85±5% of total Rad21 protein levels were depleted (3 independent experiments, $p<0.05$ assesed by Wilcoxon Rank Sum and Signed Rank Test). **d)** Maximum projections of immunocytochemistry experiments and confocal imaging of ASTs *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ after 96 hours of infection with Cre-Ad and stained for Rad21 (red) and actin (green). DNA has been counterstained with DAPI (blue) (Scale bar=20µm). Rad21 intensity from confocal images was quantified by Volocity software and is represented as a violin plot. Median values were as follows: $6x10^6$ (*Rad21*$^{Lox/Lox}$), $1.9x10^6$ (*Rad21*$^{Δ/Δ}$) arbitrary units (N_{nuclei}=350 from 3 independent experiments). Statistical significance was assessed using a Wilcoxon Rank Sum and Signed Rank Test ($p<2.2x10^{-16}$).

COHESIN-DEFICIENT ASTS SHOW INCREASED NUCLEAR VOLUMES AND MORPHOLOGICAL CHANGES

Given that previous reports had shown that cells from CdLS patients, or cells expressing mutations in cohesin subunits or the cohesin loading factor, exhibit an increase in nuclear volume, local de-compaction of chromatin and disruption of subnuclear structures without showing cohesion defects or cell cycle changes (Gard et al., 2009; Nolen et al., 2013), cohesin-deficient ASTs were checked for similar nuclear morphology changes.

Confocal microscopy was used to assess if cohesin-deficient post-mitotic ASTs experienced nuclear changes. First, 3-dimensional images of the nuclei of $Rad21^{WT/WT}$, $Rad21^{WT/WT}$+Ad-Cre, $Rad21^{Lox/Lox}$ and $Rad21^{\Delta/\Delta}$ ASTs using DAPI staining were collected in order to estimate nuclear volume ('Find object' feature of Volocity software) (**Figure 15a,b,** microscopy and analysis described in CHAPTER 5). This analysis revealed an 18% increase in nuclear volume in $Rad21^{\Delta/\Delta}$ ASTs compared to $Rad21^{Lox/Lox}$ controls (Nnuclei=350 from 3 independent experiments, $p=1.3x10^{-10}$ assessed by Wilcoxon Rank Sum and Signed Rank Test), which was a statistically different change compared to the Cre-Ad control alone (8% increase in $Rad21^{WT/WT}$+Ad-Cre compared to $Rad21^{WT/WT}$ ASTs, Nnuclei=229 from 3 independent experiments, p=0.04 assessed by Wilcoxon Rank Sum and Signed Rank Test) (**Figure 15a,b**). Rad21 intensity was quantified in the same nuclei and showed a 75% decrease in Rad21 intensity in $Rad21^{\Delta/\Delta}$ ASTs compared to $Rad21^{Lox/Lox}$ controls (Nnuclei=350 from 3 independent experiments, $p<2.2x10^{-16}$ assessed by Wilcoxon Rank Sum and Signed Rank Test), which was not observed in the Cre-Ad control alone ($Rad21^{WT/WT}$+Ad-Cre compared to $Rad21^{WT/WT}$ ASTs, Nnuclei=229 from 3 independent experiments, p=0.075) (**Figure 15a,b**). Therefore, $Rad21^{\Delta/\Delta}$ ASTs compared to their controls have lost the majority of Rad21 protein to a similar extent as that previously shown by western blot (**Figure 14**), which correlated with a significant increase in nuclear volume.

Closer examination of cohesin-deficient ASTs indicated that many nuclei were misshapen or exhibited abnormal morphologies, which were not as present in control cells (**Figure 15c**). However, DAPI staining with respect to the heterochromatin regions was similar to the $Rad21^{\Delta/\Delta}$ ASTs compared to $Rad21^{Lox/Lox}$ controls. Based on only visual inspection, nuclei were scored as normal (N) or abnormal (AN) shapes and the examples of each category are shown in **Figure 15c**. Twice as many abnormal nuclei were found in $Rad21^{\Delta/\Delta}$ ASTs than in $Rad21^{Lox/Lox}$ (20% and 8%, respectively,

Nnuclei=120 from 3 independent experiments) (**Figure 15c)**. Altogether, this data indicated that cohesin-deficient post-mitotic ASTs underwent a significant change in nuclear volume and exhibit abnormal nuclear morphologies.

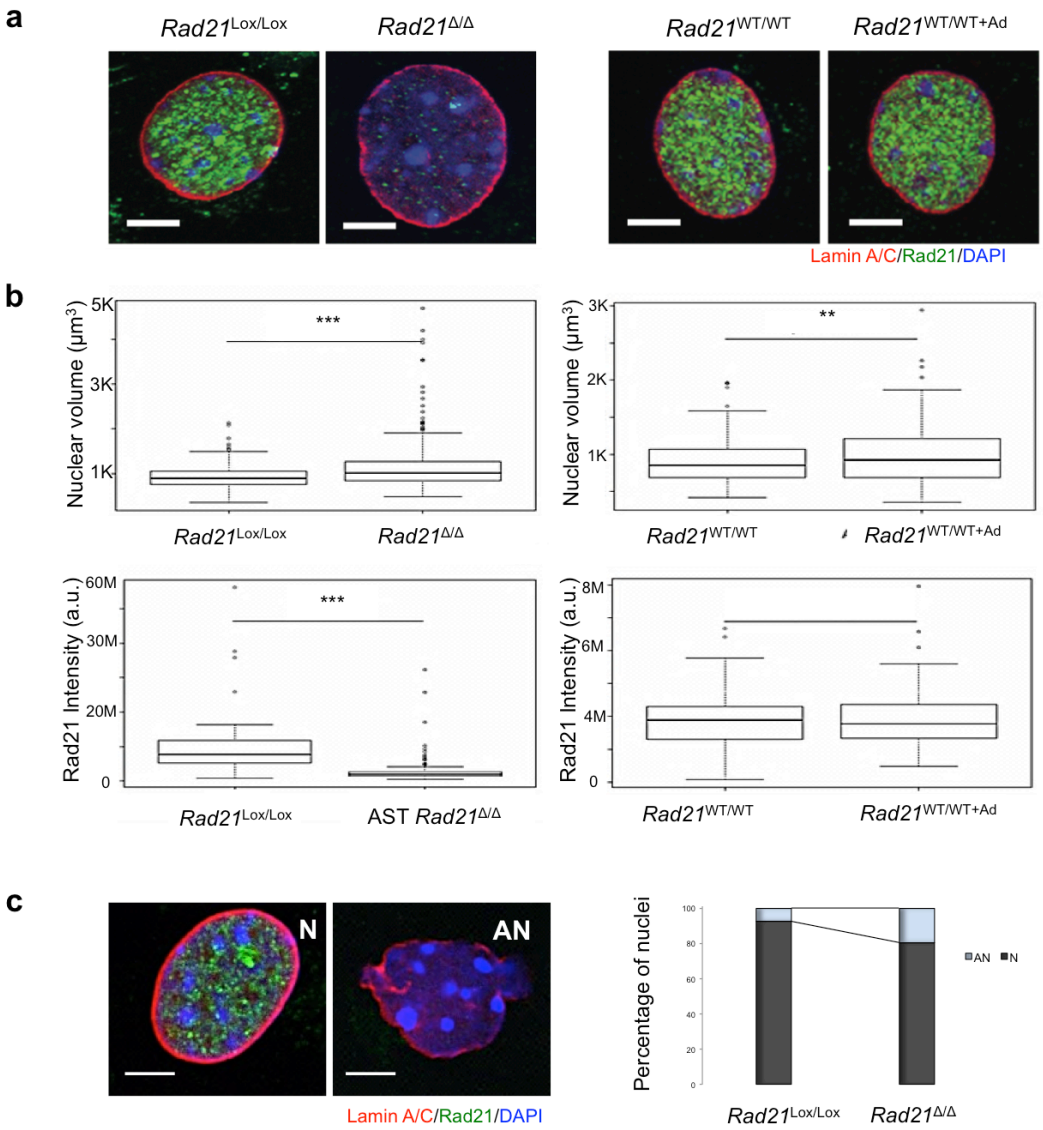

Figure 15.Cohesin-deficient ASTs have increased nuclear volumes and morphological changes. a) Maximum projections of immunocytochemistry experiments and confocal imaging of AST *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ nuclei as well as AST *Rad21*$^{WT/WT}$ and *Rad21*$^{WT/WT +Ad}$ stained for Rad21 (green) and Lamin A/C (red). DNA was counterstained with DAPI (blue) (Scale bar=5µm). **b)** Whiskers and boxes indicate all and 50% of values respectively of nuclear volume and Rad21 intensity measurements using the 'find object' feature of Volocity and they were as follows: 890 and 7.6x10^6 (*Lox/Lox*), 1008 and 1.9x10^6 (*Δ/Δ*), 853 and 3.7x10^6 (*WT/WT*), 924 and 3.7x10^6 (*WT/WT+Ad*) (N_{nuclei}=350 and 226 from 3 independent experiments) in µm^3 and arbitrary units. Statistical significance was assessed by Wilcoxon Rank Sum and Signed Rank Test (volume and intensity *Lox/Lox*: p=1.3x10^{-10}, p<2.2x10^{-16}; *WT/WT*: p=0.04, p=0.75, respectively). **c)** Single plane from confocal images and nuclear morphology was scored as normal (N) or abnormal (AN). Graphical representation of the percentages of nuclei found N or AN in AST *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ cultures, 8% and 20% AN respectively (N_{nuclei}=120 from 3 independent experiments) (Scale bar=5µm).

COHESIN-DEFICIENT CELLS EXHIBIT GLOBAL CHANGES IN GENE REGULATION

The role of cohesin in regulating gene expression at several loci in the genome (Dorsett et al., 2005; Hadjur et al., 2009; Horsfield et al., 2007; Pauli et al., 2008; Wendt et al., 2008) taken together with the finding that cohesin-deficient cells exhibited changes in nuclear size and shape, prompted me to investigate the possibility that the latter were accompanied by changes in gene regulation genome-wide in cohesin-deficient cells.

Gene expression was studied using RNA-Seq methods with the aim to comprehensively analyse the changes in the global transcriptome upon deletion of cohesin in post-mitotic cells. The Illumina Tru-Seq protocol was used, enriching specifically for polyA+ mRNA species and allowing for the analysis of known genes. Briefly, the protocol involved mRNA extraction followed by DNase treatment to fragment the mRNA and enrichment for mRNA species (poly-A+ in this case) and finally cDNA library preparation following cluster generation and sequencing. Libraries were sequenced in a strand-specific manner on the Genome Analyzer IIx by Illumina Inc. The sequencing reads were processed by Dr Wen-Ching Chan (a bioinformatian in the Hadjur lab). TopHat was used for the alignment of the reads to the mouse genome (mm9) and Cufflinks for the assembly of the transcripts (Trapnell et al., 2010). Transcript expression is represented by a value, called FPKM (Fragments Per Kilobase of transcript per Million mapped reads). RNA samples were prepared from both *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ ASTs. RNA was collected in replicate from two time-points (72 and 96 hours) post Cre-Ad addition. These samples were used to build confidence in the genes which were called differentially expressed since true cohesin gene targets. RNA was confirmed to be of good quality before proceeding. Only samples with an RNA integrity number (RIN) of 8 or above were processed further in both NSCs and ASTs (**Figure 16a, 17a**). RIN is a tool, which allows the evaluation of RNA integrity by considering the entire electrophoretic trace of the RNA sample. This includes the presence or absence of degradation products. In this way, interpretation of an electropherogram is facilitated, comparison of samples is enabled and repeatability of experiments is ensured. The assigned RIN is independent of sample concentration, instrument and analyst therefore becoming a de facto standard for RNA integrity. qRT-PCR and Western blot for the detection of Rad21 mRNA (2 independent experiments) and protein levels confirmed that Rad21 was depleted in both *Rad21*$^{\Delta/\Delta}$ NSCs and ASTs (**Figure 16b,c 17b,c**).

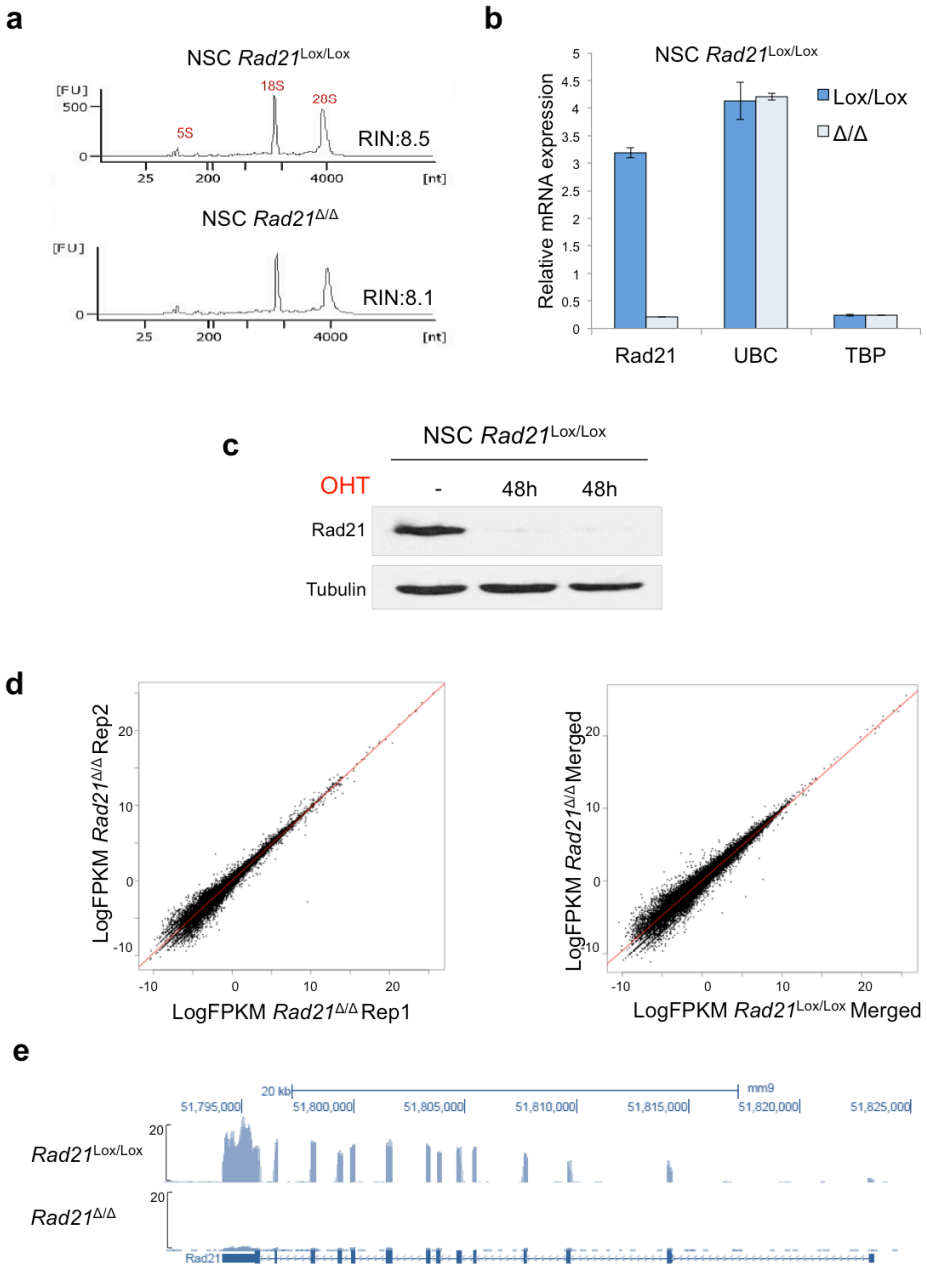

Figure 16. RNA-Seq libraries and transcriptional profiles from Rad21-deficient NSCs. **a)** RNA quality was assessed using the Agilent Bioanalyser. Electropherograms of $^{Lox/Lox}$ and $^{\Delta/\Delta}$ NSCs revealed that RIN values were all above 8. **b)** Relative mRNA levels by qPCR analysis of *Rad21* and housekeeping genes (*Tbp, Ubc*) in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ NSCs (2 independent experiments). **c)** Rad21 protein expression levels by Western blot for Rad21 and tubulin probed on the same membrane in the samples above. **d)** Scatter plots represent the FPKM values in the logarithmic scale (Log_2FPKM) of NSCs *Rad21*$^{\Delta/\Delta}$ replicate 1 versus replicate 2 (Pearson's correlation coefficient rNSC=0.9890, left plot) and NSCs *Rad21*$^{Lox/Lox}$ versus *Rad21*$^{\Delta/\Delta}$ (rNSC=0.9890, right plot). **e)** Representation of RNA-Seq reads aligned at the Rad21 gene locus in mouse genome (mm9) as they are represented in the UCSC Genome Browser in NSCs *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$.

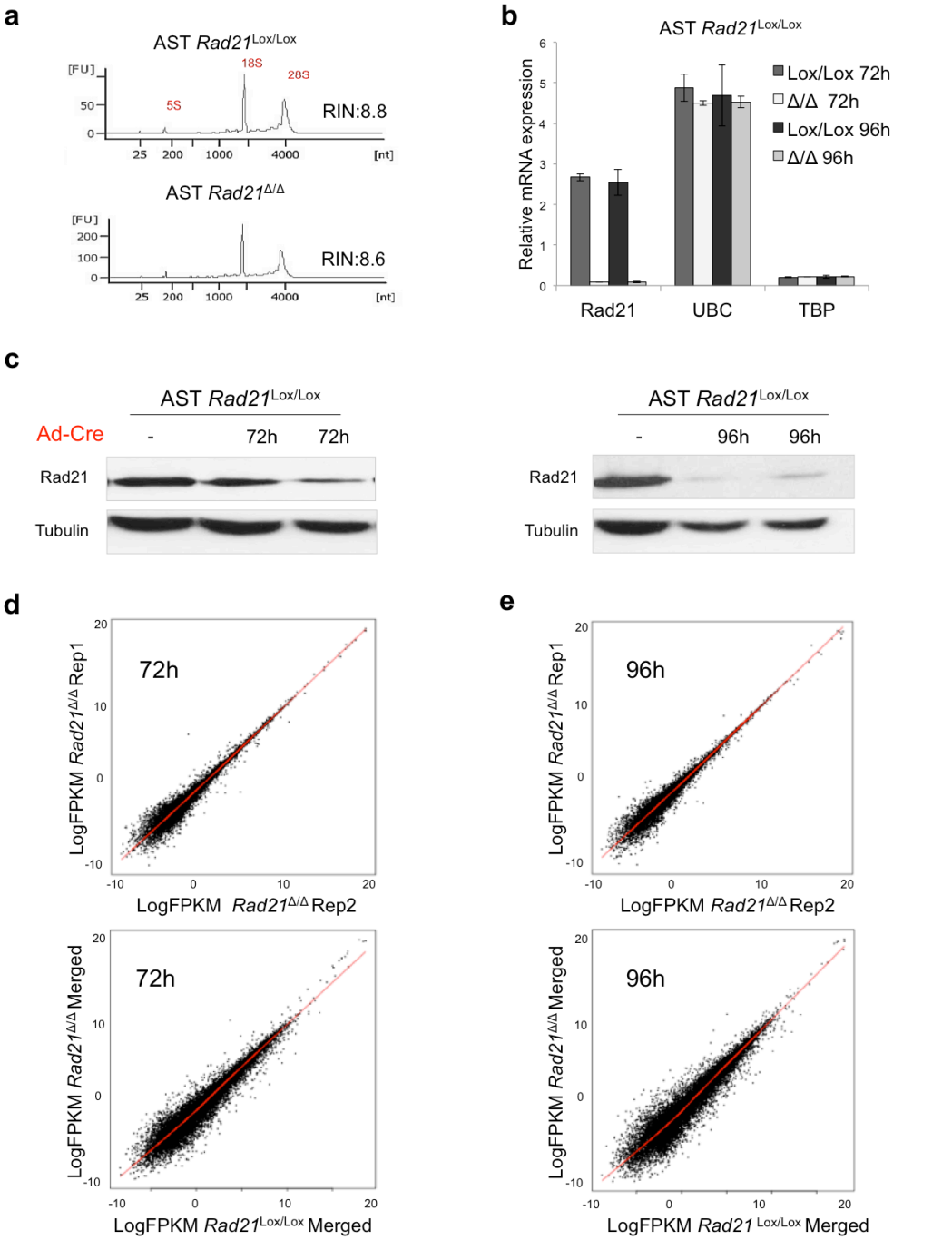

Figure 17.RNA-Seq libraries and transcriptional profiles from Rad21-deficient ASTs. **a)** RNA quality was assessed using the Agilent Bioanalyser. Electropherograms of *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ AST revealed that RIN values were all above 8. **b)** Relative mRNA levels by qPCR analysis of *Rad21* and housekeeping genes (*Tbp, Ubc*) in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ ASTs 72 hours and 96 hours post infections with Cre-Ad (2 independent experiments). **c)** Rad21 protein expression levels by Western blot for Rad21 and tubulin probed on the same membrane in the samples above. **d,e)** Scatter plots represent the FPKM values in the logarithmic scale (Log_2FPKM) of ASTs *Rad21*$^{Lox/Lox}$ replicate 1 versus replicate 2 (Pearson's correlation coefficient $rAST_{72h}$=0.9899, $rAST_{96h}$=0.9908, upper plots) and ASTs *Rad21*$^{Lox/Lox}$ versus *Rad21*$^{\Delta/\Delta}$ ($rAST_{72h}$=0.9694, $rAST_{96h}$=0.958, lower plots) in both timepoints.

In ASTs *Rad21*$^{\Delta/\Delta}$ at 72 and 96 hours, Rad21 mRNA expression was detected at similar levels with a percentage of 5% Rad21 expression remaining (**Figure 16b 17b**), whereas the decreasing levels of Rad21 protein was achieved in the two timepoints (**Figure 17c**). Thus, not only did I generate Rad21-deficient NSCs and ASTs but also two timepoints of decreasing Rad21 protein levels in post-mitotic ASTs to investigate transcriptional changes in Rad21 decreasing dosage.

Cohesin-deficient NSCs and ASTs exhibit transcriptional changes. A scatter plot of FPKM values revealed that there was a greater correlation between the NSCs as well as ASTs *Rad21*$^{\Delta/\Delta}$ replicates 1 and 2 (Pearson's correlation coefficient r_{NSC}=0.9890, r_{AST}=0.9880) compared to *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ (Pearson's correlation coefficient for merged replicates r_{NSC}=0.9791, r_{AST}=0.9554) (**Figure 16d, 17d,e, 18-19**).

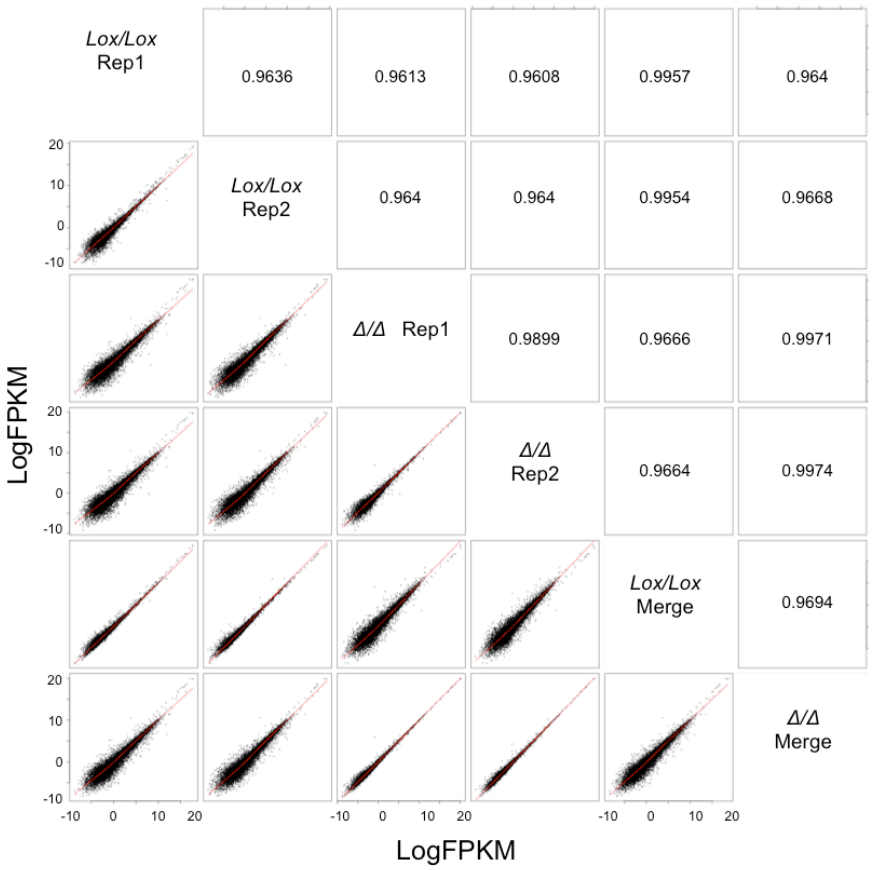

Figure 18. Transcriptional profiles in Rad21-deficient ASTs at 72 post-infections. Scatter plots represent the FPKM values in the logarithmic scale (Log_2FPKM) of ASTs *Rad21*$^{Lox/Lox}$ 72 hours post-infections with Cre-Ad all replicates and merged versus *Rad21*$^{\Delta/\Delta}$ all replicates and merged. (Pearson's correlation coefficient represented in the squares that the two samples in comparison are crossing).

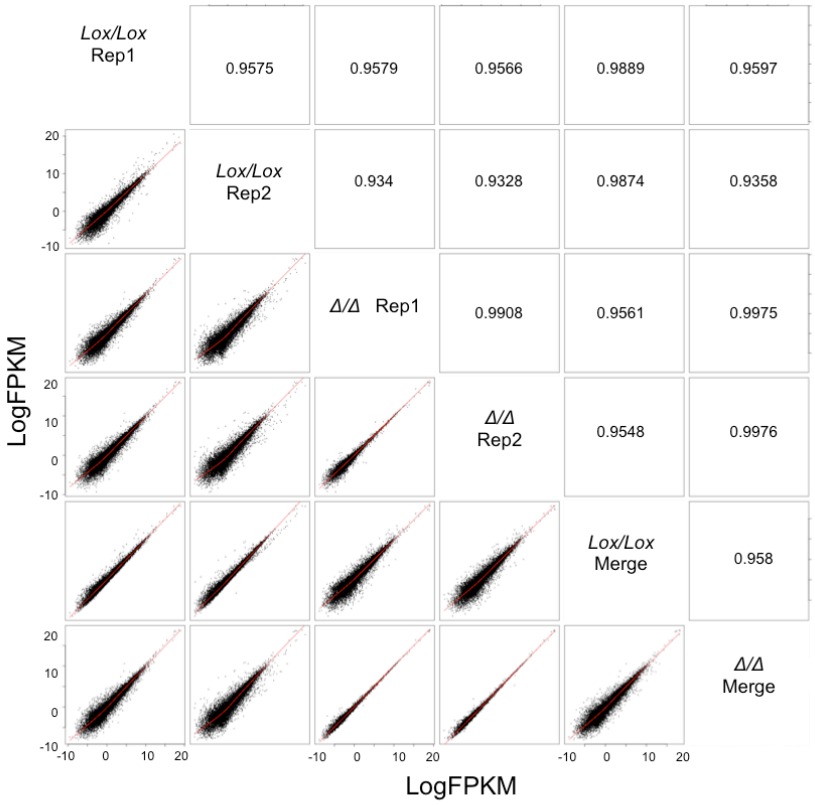

Figure 19. Transcriptional profiles in Rad21-deficient ASTs at 96 hours post-infections. Scatter plots represent the FPKM values in the logarithmic scale of ASTs *Rad21*$^{Lox/Lox}$ 96 hours post-infections with Cre-Ad all replicates and merged versus *Rad21*$^{\Delta/\Delta}$ all replicates and merged. (Pearson's correlation coefficient represented in the squares that the two samples in comparison are crossing).

By showing that the replicates of each sample were more correlated than the control sample with its respective knockdown sample, it was shown that changes in transcription occurred in NSCs and ASTs, respectively, due to Rad21 loss. Interestingly, the comparison between all of the AST timepoints (**Figure 20**) revealed that *Rad21*$^{Lox/Lox}$ ASTs in the two timepoints were less correlated (r_{AST}=0.967) than their replicates, implying ongoing transcriptional changes occurring during ASTs differentiation by BMP4 treatment between 72 and 96 hours but not as pronounced by these caused by Rad21 loss.

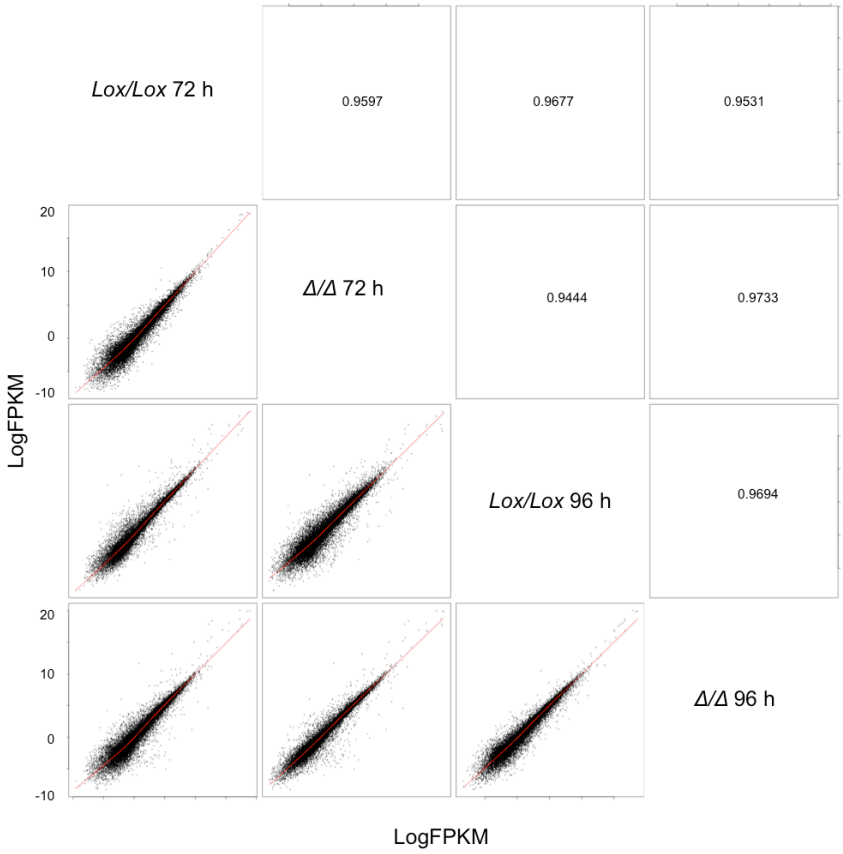

Figure 20. Transcriptional de-regulation persists in Rad21-deficient ASTs between 72 and 96 hours post-infections. Scatter plots represent the FPKM values in the logarithmic scale (Log_2FPKM) of ASTs *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ 72 hours merged versus *Rad21*$^{\Delta/\Delta}$ 96 hours merged post-infections with Cre-Ad (Pearson's correlation coefficient represented in the squares that the two samples in comparison are crossing).

Initially, to identify the set of genes which are differentially expressed (DE), a statistical method was employed by Dr W.C. Chan, where a z-score was calculated for each FPKM value indicating the distance of that value from the mean in units of standard deviations and a z-score of more than 1.5 was set as the threshold. Using this method for all identified transcripts, 2,798 genes were identified as being DE in AST *Rad21*$^{\Delta/\Delta}$, representing an 11.3% of the genes expressed (**Figure 21a**, right scatter plot) with an average fold change of 4. This method allowed for the identification of DE genes with a wide range of expression values, including lowly expressed genes. Differentially expressed genes were found to be both up- (4.9%) and down-regulated (6.4%) (**Figure 21b**). This observation was in agreement with other published work in mice and flies (Dorsett et al., 2005; Pauli et al., 2008; Remeseiro et al., 2012b), where also minor fold-changes in gene expression were observed.

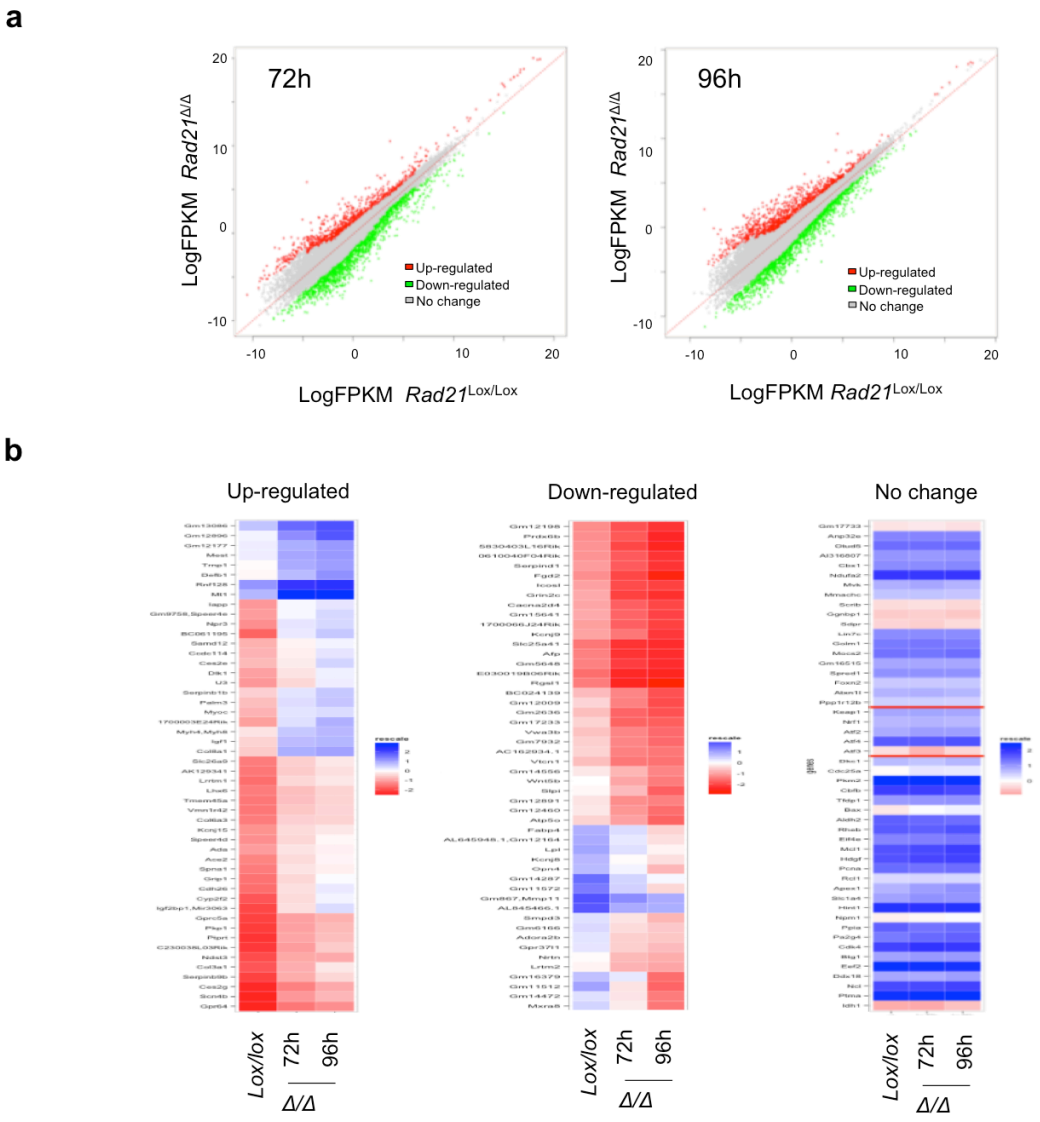

Figure 21. Cohesin-deficient ASTs exhibit robust changes in gene expression. a) Scatter plots represent the FPKM values in the logarithmic scale of AST *Rad21*$^{Lox/Lox}$ versus *Rad21*$^{Δ/Δ}$ at 72 hours (left) and 96 hours (right) post-infection using z-score method indicating the differentially expressed in both timepoints, up-regulated (red) and down-regulated (green). **b)** Heatmaps representing in colour code the changes in FPKM values of 50 genes in three separate classes (up-, down-regulated, non-changing) using the z-score method in AST *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ at 72 and 96 hours post-infections. Higher and lower expression values are represented in bleu and red. The genes and their expression values (FPKM) are presented in Table 5 in the APPENDIX.

Interestingly, AST *Rad21*$^{Δ/Δ}$ in the two timepoints were shown to be highly correlated (Pearson's correlation coefficient, r_{AST}=0.9733) (**Figure 20**). Applying the same thresholds as described above to the consecutive timepoints revealed that set of genes found de-regulated at 72 hours [10.2% of expressed genes, up- (3.6%) and down-regulated (6.6%)] is very similar to the set of genes at 96 hours (1,190 overlapping DE genes) and they followed the same trend of up- or down-regulation in the two

consecutive timepoints (**Figure 21a,b**). The above results support the idea that while the changes in transcription are subtle, they do not represent transcriptional noise and in fact represent real transcriptional changes due to Rad21 knockdown.

To investigate potential hotspots or regions in the genome where cohesin is regulating transcription, the locations of the transcription start sites of either the up-regulated (red), down-regulated (green) or non-changing (grey) genes from ASTs simultaneously were represented in the respective colours on all chromosomes (**Figure 22**). The de-regulated genes were found evenly distributed throughout the genome indicating that there were no obvious hotspots of cohesin-dependent gene deregulation and that cohesin has a genome-wide effect on transcriptional regulation (**Figure 22**).

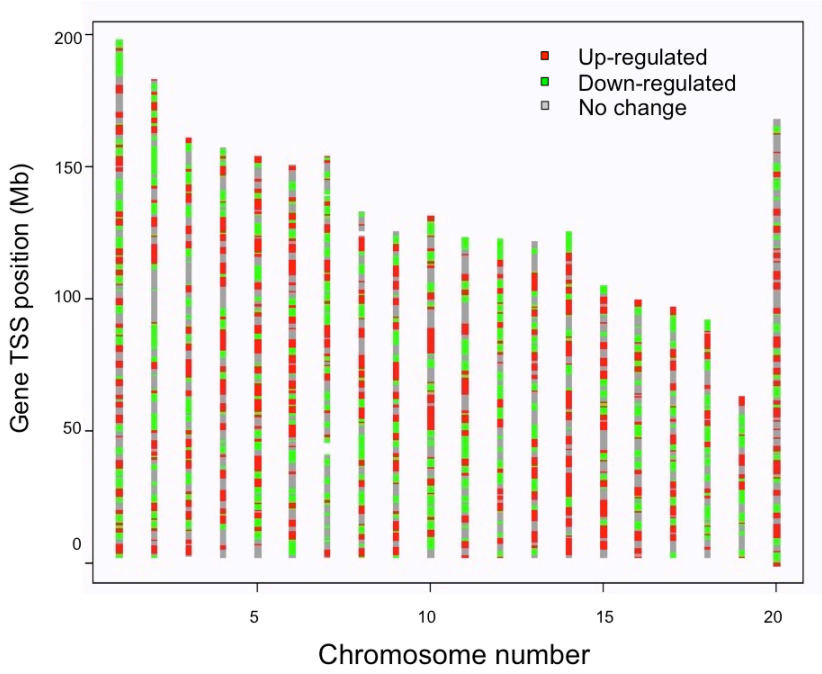

Figure 22. Cohesin-deficient ASTs exhibit global changes in gene expression. Ideogram represents the position of transcription start sites (TSS) (Mb) of the up- (red), down-regulated (green) and non-changing (grey) genes indicated on the different mouse chromosomes in AST *Rad21*$^{\Delta/\Delta}$ at 96 hours post-infections.

To determine which of the deregulated genes were directly influenced by Rad21, deregulated genes were assayed for the presence of a Rad21 ChIP peak. For this, a Rad21 ChIP-Seq analysis in NSCs and ASTs was used. Using the Model-based Analysis for ChIP-Seq (MACS) peak calling program, 18,405 and 16,156 Rad21 binding sites were identified in NSCs and ASTs respectively. MACS is a commonly used algorithm that evaluates the significance of enriched ChIP regions and improves the spatial resolution of binding sites through combining the information of both the

position of the sequencing reads and their orientation. As expected, there was a strong overlap (~70%) in Rad21 binding sites between the two cell types. Integrating the RNA-Seq and ChIP-Seq data identified that 50% of DE genes in ASTs (1,553) contained a Rad21 peak within the transcriptional unit. This is enrichment when compared to all non-changing genes containing Rad21 (4,549) (p value ≤ 0.001, Fishers Exact Test) (**Figure 23a**). While the majority of DE genes were found to be enriched for cohesin binding compared to non-DE genes, the majority of the DE genes had no cohesin at their transcriptional unit (Sofueva et al., 2013) (**Figure 23b**), implying that some of the genes which are de-regulated might arise indirectly from changes in cohesin-anchored contacts involved in mediating higher-order loops, that form their microenvironment, other than gene loops.

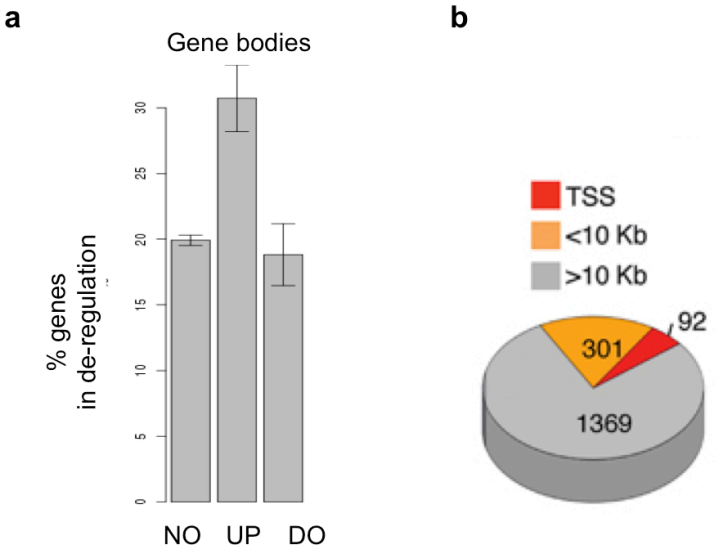

Figure 23. De-regulated genes in cohesin-deficient astrocytes are enriched for cohesin binding. a) Bargraphs represent the **p**ercentages of deregulated genes bound by cohesin at their gene bodies, in the no-, up- and down-regulated (NO, UP, DOWN, respectively) groups (p value ≤ 0.001, Fishers Exact Test). **b)** Distribution of a group of 1762 deregulated genes according to cohesin/CTCF occupancy at the transcription start site (TSS) (<1 kb), near the TSS (<10 kb) and away from TSS (>10 kb).

Altogether this analysis revealed that about 11% of transcripts were differentially expressed in cohesin-deficient ASTs, that the majority of these had small changes in gene expression and were uniformly distributed across the mouse genome. Finally, while DE genes were enriched for cohesin binding within their transcriptional unit compared to non-changing genes, the majority of DE genes did not have cohesin bound. Therefore, by integrating transcriptome analysis with cohesin binding profiles, I have revealed that cohesin's role in gene regulation involves the stabilisation of some

specific promoter-enhancers, but these may primarily stem from indirect effects of cohesin depletion (discussed below).

RAD21-MEDIATED PROMOTER-ENHANCER INTERACTIONS REGULATE THE PROTOCADHERIN A GENE CLUSTER IN COHESIN-DEFICIENT CELLS.

To discover if the genes which are sensitive to cohesin deficiency are enriched in a particular functional category, a gene enrichment analysis (pathway analysis, Panther software, p>0.05) was performed (**Figure 24**). PANTHER (Protein ANalysis Through Evolutionary Relationships) classification system is a large biological database of gene families which is used to classify and identify the function of gene product.

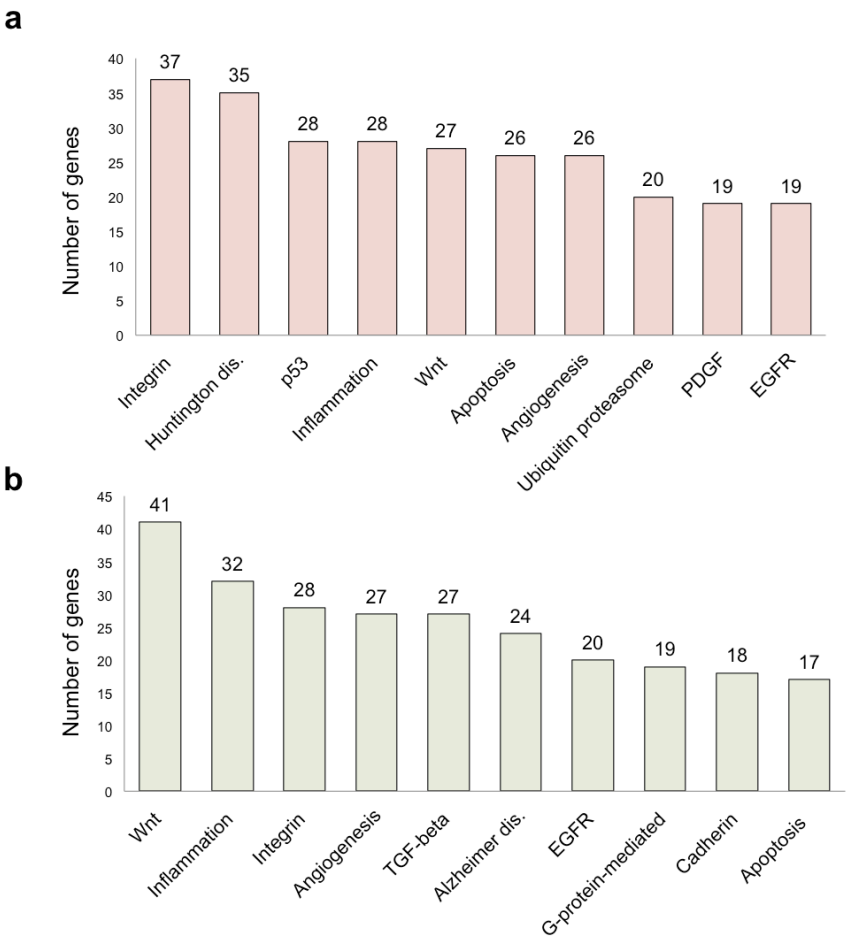

Figure 24. The deregulated pathways in cohesin-deficient ASTs. a) Plots representing the number of either up- (upper, orange) or, **b)** down-regulated (lower, green) genes involved in each of the pathways indicated in the x axis using Panther software. Numbers in the above bars represent the number of genes in each pathway (p>0.05).

Among the top ten pathways identified were genes that are involved in cell-to-cell and cell-to-cellular matrix contacts, such as the Pcdh gene cluster (**Figure 24**). The Pcdhα gene cluster is a good candidate for the investigation of cohesin role in gene regulation via direct cohesin-mediated promoter-enhancer interactions. Cohesin and CTCF have been implicated in the stochastic choice of alternative promoter usage in this region (Guo et al., 2012; Monahan et al., 2012). This cluster of genes is essential during the development and for the physiology of the nervous system (Yagi, 2008) so the neural stem cells used for this study were relevant for the investigation of Pcdh gene regulation.

The Pcdhα transcripts are transcribed from one alternative exon (α1-12) and 3 common constant exons (**Figure 25b**, UCSC Genome browser view of the Pcdh cluster of genes with Pcdhα cluster in blue). In neurons, a combination of monoallelic and biallelic expression of the Pcdhα transcripts was shown and individual neurons were suggested to express different sets of Pcdhα transcripts (Yagi, 2008). RNA-seq analysis revealed that in this particular clone of NSCs, a restricted set of Pcdhα transcripts were expressed (α4, α10, αc2) and that after Rad21 loss in either NSCs or ASTs, this specific gene set is up-regulated. This was validated using primers designed to amplify the individual alternative exons and the common constant exons, and transcript levels were assessed by qRT-PCR in *Rad21*$^{Lox/Lox}$ NSCs where Rad21 protein was depleted (**Figure 25a,c**, 3 independent experiments). Mouse liver RNA was included as a negative control, where the Pcdh cluster of genes is shown to be silent. In order to determine if the Pcdha transcripts were direct targets of Rad21, Rad21 ChIP-Seq data were used in the NSC clone which expressed Pcdhα4, α10 and αc2. Rad21 binding sites were observed at the promoters of the exons expressed. Rad21 binding sites were also mapped between constant exon 2 and 3 as well as downstream of it, where the proposed enhancer regions (i.e. HS5-1, HS7, respectively, indicated in black arrows) that regulate the expression of the Pcdhα promoters were previously mapped (Yagi, 2008). Similar results were reported for Pcdhβ/γ gene clusters and their downstream cis-regulatory elements [i.e. cluster control region, Cis Control Region (CCR) at the Diap gene located downsteam of the constant exons, indicated in black arrows]. These results agree with published studies implicating cohesin and/or CTCF in the regulation of Pcdh clusters in neural tissues (Hirayama et al., 2012; Kehayova et al., 2011). Thus, these results suggested that cohesin depletion directly led to a de-regulation of transcription in the Pcdh cluster.

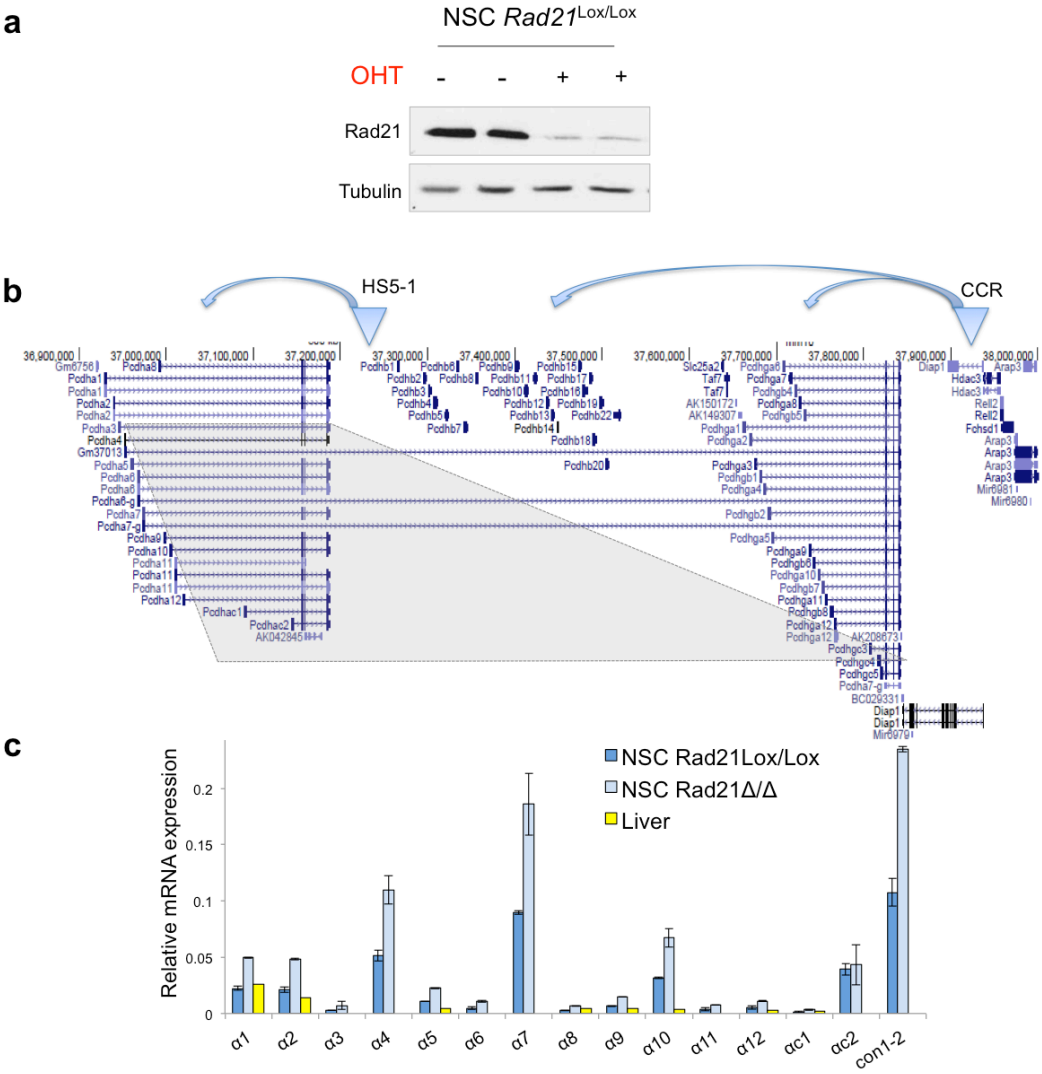

Figure 25. Protocadherin cluster of genes is de-regulated in cohesin-deficient cells. a) Rad21 protein expression levels by Western blot for Rad21 and tubulin probed on the same membrane in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ NSCs used for the qPCR analysis and 4C-Seq down-stream experiments. The two replicate samples were from the same parental line. **b)** UCSC genes in the genomic region where the Protocadherin (Pcdh) cluster of genes is located. This includes the alternative exon of Pcdhα, Pcdhβ and Pcdhγ clusters. The cis-regulatory elements of Pcdhα and Pcdhβ/γ clusters, such as HS5-1 and Cluster Control Region (CCR), respectively, are indicated in bleu arrows. **c)** Relative mRNA levels by qPCR analysis of Pcdhα alternative exons and constant exons normalised for expression of housekeeping genes (*Tbp, Ubc*) in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ NSCs as well as in liver samples, which were used here as a negative control (3 independent experiments).

In order to further investigate the role of cohesin in mediating chromatin interactions which may define the transcriptional regulation of the Pcdh cluster, high-resolution 4C-Seq experiments were performed. 3C techniques have been developed to study genome organisation in higher resolution and output than by microscopy methods. The premise behind 3C is that when chromatin fragments are 'proximal' in nuclear space, they will be more likely to be ligated together and could represent interactions between

chromatin fragments. A derivative of 3C is 4C-Seq, which captures chromatin interactions between one known locus (i.e. bait) and many other possible interactions (van de Werken et al., 2012). This technique was optimised for both ASTs and NSCs.

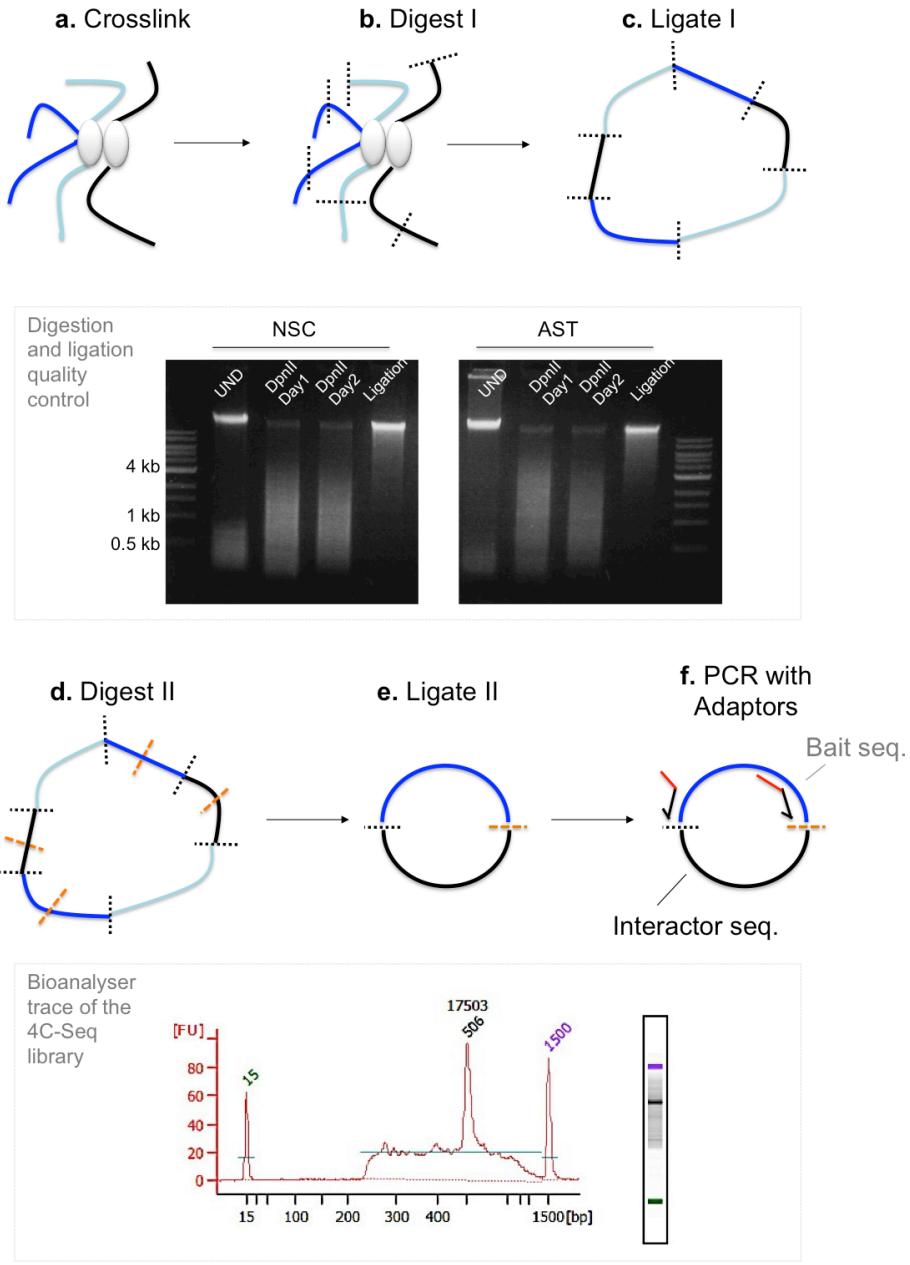

Figure 26. 4C-Seq protocol and quality control steps in mouse NSCs and ASTs. Main steps of the protocol were as follows: **a)** Crosslinking proteins with DNA using 1% folmadahyde (20, 30 minutes) for NSCs and ASTs, respectively); **b,c,d,e)** Two rounds of digestion of DNA fragments (DpnII and Csp6I) and ligation to decrease the size of the fragments and increase resolution of the method; **f)** PCR reaction for long-template DNA (Roche) with primers recognising the bait sequence and containing the Illumina adapter (red) followed by sequencing. Libraries contain the adapter, the recognition site and the interactor fragment sequences so that the libraries with different baits were multiplexed before sequencing. Adapted from van de Werken *et al*. (2012). UND: undigested sample, DpnII Day 1 and Day 2: digested samples with DpnII for 1 and 2 days, respectively.

By using two consecutive rounds of digestion (with DpnII and Csp6I enzymes) and ligations (**Figure 26a-e**), the pool of possible interacting fragments becomes very complex and thus higher confidence interactions can be identified. PCR is performed using 'bait' primers at the restriction sites of the bait sequence (**Figure 26f**, in black). Therefore, the DNA fragments captured in close proximity with the bait sequence in the nucleus will be amplified and the frequency of these interactions will be detected (**Figure 26f**).

Here, the baits were designed to be located at cohesin (i.e. Rad21) binding sites mapped by ChIP-Seq (discussed in the previous paragraph) at expressed promoters (i.e. Pcdhα10) or at known enhancer regions (i.e. HS5-1 or CCR, indicated in bleu arrows), (**Figure 27**, top ChIP-Seq panels) in both *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ NSCs. Interestingly, 4C-Seq results show that HS5-1 enhancer region preferentially interacts with the promoters which were expressed in *Rad21*$^{Lox/Lox}$ cells (**Figure 27**, middle 4C-Seq panels). Reciprocal contacts were also detected between one of the expressed promoters, namely Pcdhα10, and the HS5-1 enhancer region in *Rad21*$^{Lox/Lox}$ cells. These results show the existence of specific chromatin contacts between the expressed promoters and regulatory elelements *in cis*. Upon cohesin depletion, a decrease in the specific chromatin contacts between HS5-1 enhancer region and the expressed promoters was reported, implying a role from cohesin in mediating these loops. Similar results were obtained from investigating an additional enhancer region CCR, which specifically regulates the Pcdhβ-γ clusters and not the Pcdhα. The CCR region was preferentially interacting with the Pcdhβ-γ promoters, which is known to regulate, and not with the Pcdhα. Upon Rad21 depletion, these specific chromatin contacts were also decreased. Thus, cohesin depletion resulted in a loss of long-range contacts anchoring distal enhancers required for the expression of genes within the Pcdh cluster, which were shown to be de-regulated in cohesin-deficient cells.

Dr S. Sofueva had established the Hi-C method in the lab to characterise cohesin-anchored chromatin loops (Sofueva et al., 2013). Hi-C is a derivative of the 4C approach described above, but instead of measuring long-range interactions from specific viewpoints, it provides a catalogue of all chromatin contacts occurring simultaneously and thus leads to an unbiased global analysis of chromatin structure (Lieberman-Aiden et al., 2009). Results from a Hi-C experiment are represented in Hi-C contact maps, such as in **Figure 27**, where the normalised frequencies (colour scale in log) of chromosomal interactions are represented as a matrix with the diagonal representing this region in the linear chromosome distance. The Hi-C maps reveal

regions of high-contact frequency separated by distinct 'insulation' points. Such structures had been termed 'chromosomal domains' or TADs and the insulation points are termed TAD borders (**Figure 27**).

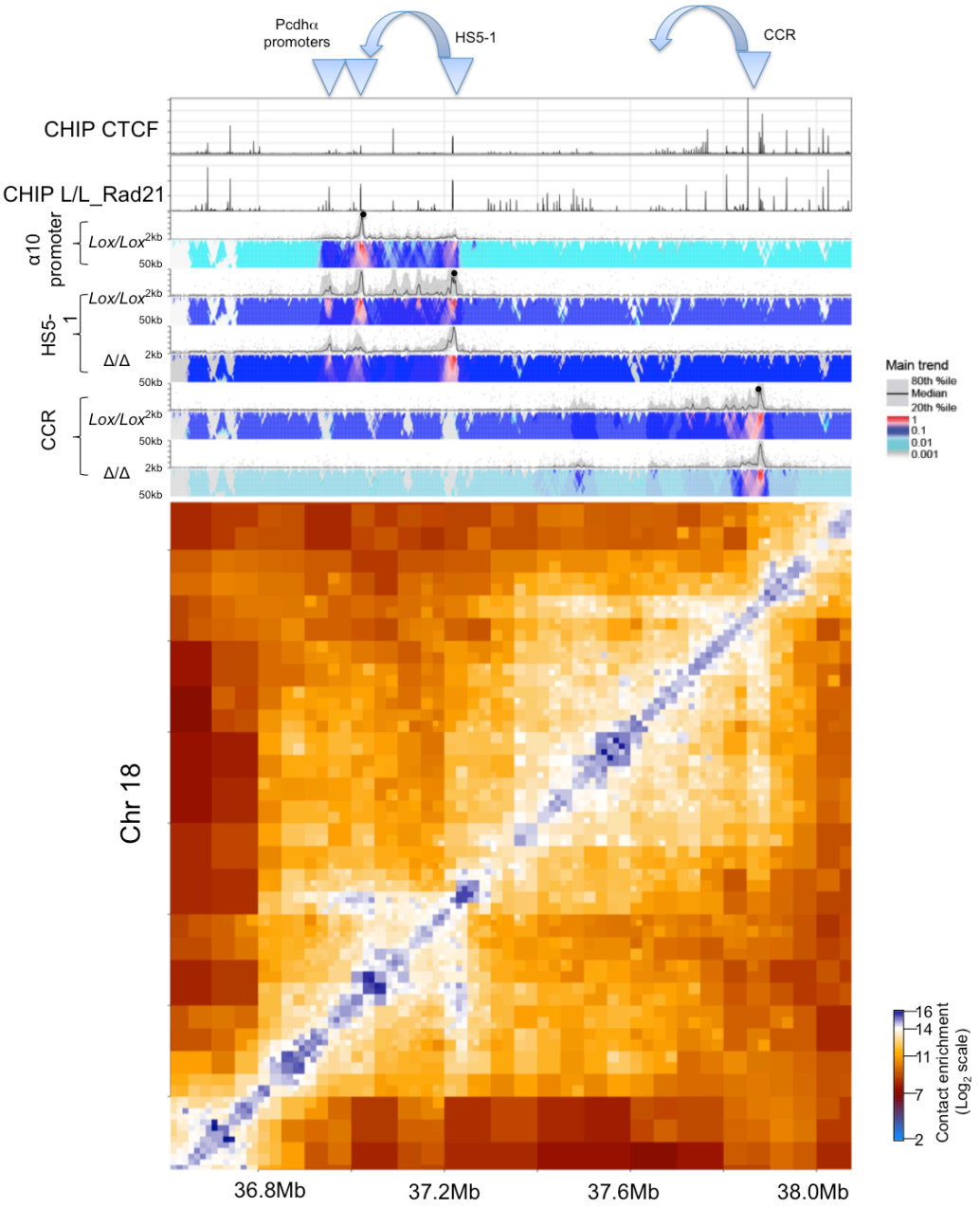

Figure 27. Cohesin-anchored chromatin contacts shape the regulatory landscape in the Protocadherin cluster. HiC contact matrix of a 2 Mb region on chromosome 18 (lower) where the Pcdh gene clusters are located in *Rad21*$^{Lox/Lox}$ NSCs highlighting the topological associated domain encompassing the Pcdhα cluster and Pcdhβ,γ (grey). ChIP-Seq signal aligned was for mouse cortex and *Rad21*$^{Lox/Lox}$, respectively (upper). Results from 4C-seq in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ NSCs for baits (indicated in black dots) designed to Rad21 binding site at the promoters of Pcdhα10, at the HS5-1 enhancer and the Cis Control region (CCR). Each 4C-seq experiment is represented by the median normalised coverage in a sliding window of 5 kb (top) and a multi-scale domainogram indicating normalised mean coverage (windows range 2-50 kb).

I examined the chromosomal structure, determined by Hi-C around the region containing the Pcdh gene cluster on chromosome 18. The Pcdhα cluster, together with its cis-regulatory elements were found to be located within the same Hi-C domain (**Figure 27**). The Pcdhα and Pcdhβ/γ clusters are located in two separate Hi-C domains (**Figure 27**). Therefore, these findings suggested that cohesin-mediated gene-loops may be contained within larger chromatin domains, and that chromosomal domain structure may be important for correct gene regulation.

Hence, Pcdhα and Pcdhβ/γ clusters were found structurally contained within two adjacent Hi-C chromatin domains, which also enclosed their respective cis-regulatory regions. In Pcdhα, I showed that cohesin contacts mediated promoters and enhancers interactions which in turn regulate its expression. This is a characteristic example of cohesin-mediated gene regulation via maintaining promoter-enhancer interactions *in cis*.

COHESIN-MEDIATED CHROMATIN INTERACTIONS ORGANISE CHROMOSOMAL DOMAIN ARCHITECTURE

The data described above showed that *Rad21*$^{\Delta/\Delta}$ NSCs and ASTs experienced transcriptional de-regulation. These results were in line with previously published work describing cohesins effect on gene regulation (Dorsett et al., 2005; Pauli et al., 2008; Wendt and Peters, 2009). One mechanism by which cohesin can regulate gene expression is by anchoring long-range chromatin loops which then physically connect cis-regulatory regions to target loci (Hadjur et al., 2009; Kagey et al., 2010; Nativio et al., 2009; Seitan et al., 2011) as well as at the Pcdh cluster shown here. We reasoned that cohesin might anchor chromatin loops from CTCF binding sites not just at specific genes, but also throughout the genome. To address this, Dr S. Sofueva prepared Hi-C libraries from *Rad21*$^{\Delta/\Delta}$ NSCs and ASTs to dissect the role of cohesin/CTCF in mediating genome-wide chromosomal contacts published in (Sofueva et al., 2013). This work revealed that cohesin-anchored chromatin contacts are involved in shaping domain structure at different size scales and that cohesin depletion was quantitatively correlated with chromatin domain decompaction.

A particular interesting example which revealed cohesin's roles in chromosome organisation and how this organisation could impact gene expression came from the mTOR inhibitor Deptor, and its neighboring gene Col14a1 (**Figure 28**). The analysis of

the RNA-seq datasets in replicates revealed that Deptor is highly expressed and becomes 2.6-fold downregulated, while Col14a1 is silenced and becomes 7-fold upregulated in Rad21-deficient ASTs (**Figure 28a**, the RNA-Seq data described in CHAPTER 2). Cohesin binding sites are localised to the promoter of the Deptor and Col14a1 genes, as well as at sites ~15 kb upstream of Deptor gene. From the Hi-C contact maps, it appeared that Deptor and Col14a1 had a 'nested loop structure' within a larger chromosomal domain. In Rad21-deficient cells, the domain becomes decompacted however the overall structure of the domain is retained. To dissect the effect of cohesin loss in this region, a timecourse 4C-seq experiment was performed from cohesin sites which were at either the Deptor promoter or upstream of it demarcating the Hi-C domain enclosing the gene. Rad21 protein levels were quantified by western blot and a progressive loss of Rad21 protein levels was observed over time (**Figure 28b**). I prepared 4C-Seq libraries from each timepoint corresponding and 4C baits were assayed for their contact profile during the progressive loss of Rad21 (**Figure 28c**). The results showed the progressive loss of cohesin-mediated chromatin interaction is accompanied with the one in cohesin protein levels at this region either from the promoter or upstream of that viewpoints. Interestingly, it also revealed that cohesin-anchored contacts mediating gene-loops (i.e. *Deptor*) were lost quicker than the domain-forming loops shown in several regions [(Sofueva et al., 2013) and unpublished work arising from this thesis)] (**Figure 28d**).

Taken together, these findings proposed that cohesin is a key player in genome organisation whether this involved creating 'gene-loops' or longer-range 'gene-free loops'. However, while loss of cohesin protein was accompanied with a loss of chromatin contacts in the 4C-Seq regions, it was suggested that cohesin-anchored contacts at the gene-free loops were more resistant to cohesion depletion.

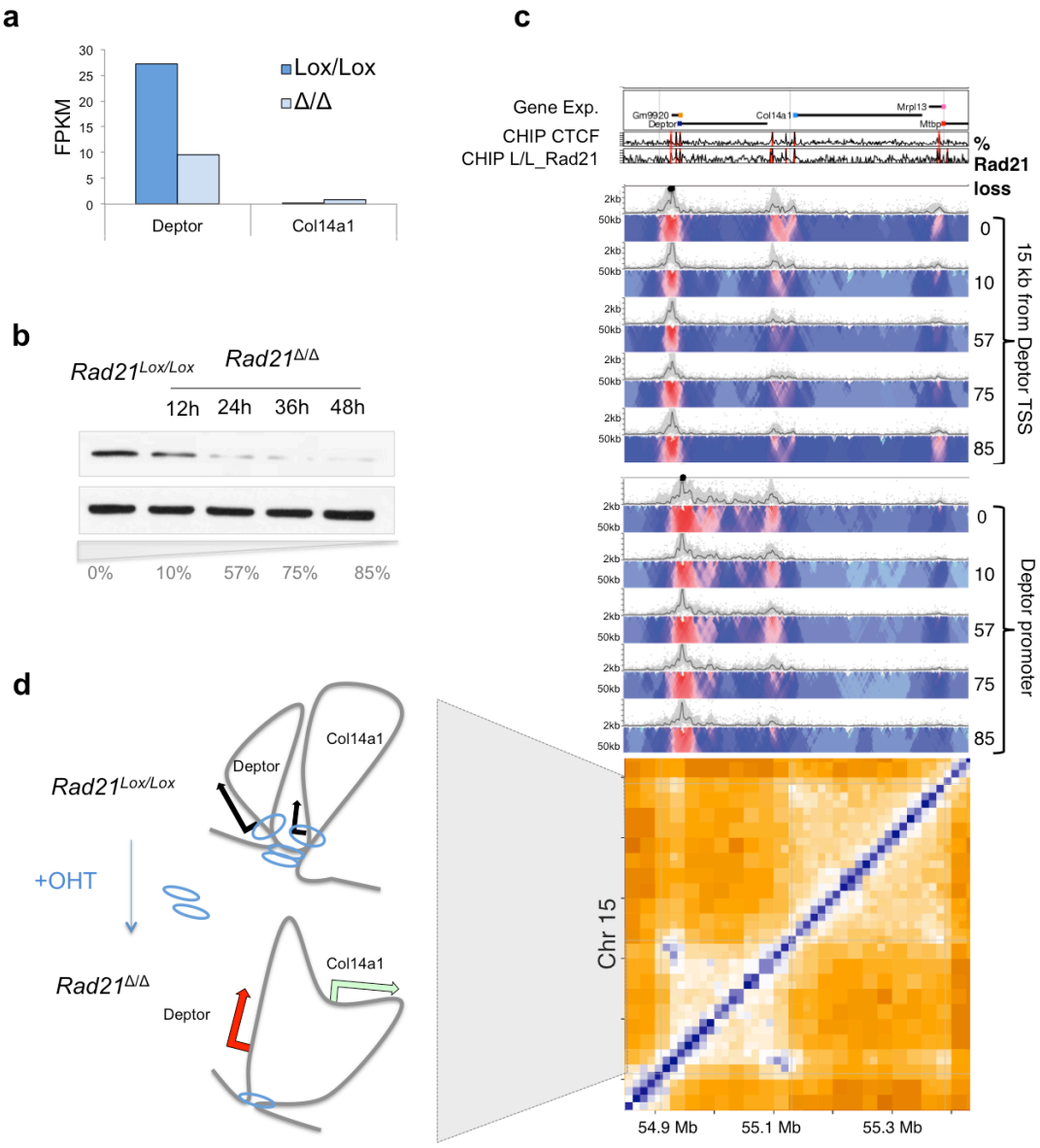

Figure 28. The loss of cohesin-anchored chromatin contacts, either at gene promoters or gene-free loops, are accompanied by the loss in cohesin protein levels. a) mRNA expression levels of *Deptor* and *Col14a1* expressed as FPKM in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ ASTs. **b)** Western blot analysis for Rad21 and tubulin proteins during OHT timecourse in *Rad21*$^{Lox/Lox}$ NSCs. Quantification of Rad21 intensity by ImageJ software, normalised to tubulin levels in each timepoint and percentage of Rad21 loss indicated (grey). **c)** 4C-seq results in a timecourse of Rad21 depletion in *Rad21*$^{Lox/Lox}$ NSCs for baits (black dots) designed to Rad21 binding site at the promoter of *Deptor* and at 15 kb upstream of its transcription start site (TSS). Shown are the % loss in Rad21 protein levels at each timepoint based on a quantitative Western blot analysis. Each 4C-seq experiment is represented by the median normalised coverage in a sliding window of 5 kb (top) and a multi-scale domainogram indicating normalised mean coverage (windows range 2-50 kb). Also shown are the ChIP-Seq tracks for Rad21, CTCF, TSS locations and gene expression. NSCs HiC contact matrix of a 1 Mb region on chromosome 15. **d)** Graphical representation of the proposed model for Deptor region, where cohesin is at the basis of gene-loops and gene-free loops. After Rad21 depletion, both chromatin domain decompaction and gene deregulation were observed but residual chromatin contacts were still detectable.

2.3. DISCUSSION

To investigate the role of cohesin proteins and their roles in chromatin looping and gene regulation in G_1, I developed a post-mitotic cell system whereby cohesin deletion would not disrupt its essential mitotic functions. ESCs were converted *in vitro* into NSC lines, which were subsequently differentiated into AST cultures by the addition of BMP4 growth factor. The Rad21 subunit of the cohesin complex was successfully deleted, using Cre-Lox methods in post-mitotic ASTs, thus generating a robust system with which to perform downstream experiments.

Using the above system, I observed that cohesin-deficient post-mitotic cells exhibited global de-regulation of transcription, increased nuclear volume and perturbed nuclear morphology. This finding is supported by phenotypes seen in CdLS patients or cells expressing cohesin mutations (Gard et al., 2009; Nolen et al., 2013). The phenotype of perturbed nuclear size and shape can come about when cells are entering the cell cycle, or initiate cell death, which we excluded here. An additional reason why these changes take place in cohesin-deficient ASTs could be due to changes in the nuclear lamina, which is linked to genome organisation and transcription. In progeria syndromes, the mutation of one gene involved in the lamina of the nucleus (i.e. *LMNA*) is required for phenotypes such as in cohesin-deficient ASTs and this is sufficient to cause pre-mature aging in humans (Goldman et al., 2004). Hence, I propose that the phenotype in cohesin-deficient ASTs is more likely to be linked to changes in transcription and chromatin organisation.

These observations supported the decompaction of chromosomal domain structure described using Hi-C approaches. The loss in cohesin-mediated chromatin interactions often led to the de-regulation of genes flanked by cohesin sites reflecting a direct role in gene regulation via anchoring chromatin interactions between regulatory elements and promoters. However, a large number of genes did not contain any cohesin binding site within their gene body, yet were still de-regulated in $Rad21^{\Delta/\Delta}$ cells. This suggests that cohesin may indirectly regulate gene expression. I propose that this may come about cohesin loss actually disrupts the regulatory microenvironment that encompasses the affected genes. Simultaneously, the genome-wide and candidate-based 3C studies to look at genome organisation in cohesin-deficient cells showed topologically associated domain relaxation correlated with a specific loss of cohesin-mediated contact insulation. Taken together, my results support a key role for cohesin

in mediating global genome organisation and gene expression, both through direct and indirect means (Sofueva et al., 2013).

These results broadly agree with studies in non-cycling or cycling mammalian cells, with either cohesin genetic depletion (Seitan et al., 2013) or TEV-mediated cleavage (Zuin et al., 2014), respectively. The study in mouse cohesin-deficient thymocytes emphasised their analysis to cohesin transcriptional roles and proposed a role for cohesin in gene regulation internally to the TADs (Seitan et al., 2013). Zuin *et al.* using Hi-C and depletion of cohesin or CTCF via TEV-mediated cleavage or RNAi, respectively, in 293 cells reported a separate role for CTCF from its role with cohesin (Zuin et al., 2014) (discussed in the CHAPTER 4). All three Hi-C studies exploring the functional role of cohesin in global genome organiation underline an important role for cohesin and CTCF in TAD maintenance and provide further insights to their functions.

However, it was not clear yet whether cohesin and CTCF have a causal relationship with the maintenance of the TADs. While I decreased cohesin levels to about 20% of the original amount, I still detected chromatin contacts (i.e. 4C-seq) arising from cohesin sites defined by ChIP-Seq (i.e. 1.5 kb upsteam of the Deptor promoter). While our methodology (i.e. 4C-Seq) provides high-resolution, it can't exclude the possibility that factors other than cohesin and CTCF, are also present and potentially involved in maintaining chromatin loops at these sites. At the same time, recent supportive evidence from a study using CRISPR-mediated deletions of specific CTCF binding sites, propose a key role for CTCF in chromatin loop maintenance (Narendra et al., 2015). A 4C-Seq experiment at the *HOXA* locus revealed that CTCF binding sites were located at a border region separating active from inactive domains, based on epigenetic marks and transcriptional activity. Importantly, the deletion of the CTCF binding site led to the expansion of the active histone marks into an adjacent repressive domain resulting in the activation of repressed genes (Narendra et al., 2015). This was accompanied by ectopic chromatin contacts spreading up to the next CTCF binding site within that chromatin domain, implying an essential role for CTCF in the maintenance of the chromatin domain borders at this region.

The importance of CTCF as an insulator with roles in global genome organisation is also proposed from another study where a genome-wide Hi-C was integrated with ChIP-Seq studies for CTCF in liver tissues from three mammalian species (Vietri Rudan et al., 2015). This study showed that conserved CTCF sites mediate conserved chromosomal domain structure in mammalian genomes, while divergent (Species-

specific) CTCF sites mediate insulation specifically in the corresponding genome. Such divergent sites could be considered as ‘multiple CRISPR-inserted sites’ and reveal that new chromatin loops are found when a new CTCF site emerges.

In addition, both the above and Sofueva *et al.* study propose that CTCF and/or cohesin mediate chromatin structure at multiple scales and from thousands of sites across the genome. Now that it is clear that chromatin domains don’t exist only at the TAD level (1 Mb) but at multiple scales and in a hierarchical manner, these findings explain why while CTCF is enriched at TAD borders, most of the CTCF sites are located within TADs (Dixon et al., 2012; Sofueva et al., 2013). The above studies employied approaches that measure cohesin/CTCF roles at multiple scales and were able to demonstrate their functional roles in genome organisation. All these observations further support the key connection between CTCF, cohesin and chromatin looping.

CHAPTER 3.

COHESIN REGULATORS MODULATE THE STABILITY OF COHESIN AND COHESIN-ANCHORED CONTACTS IN POST-MITOTIC ASTROCYTES

CHAPTER 3. COHESIN REGULATORS MODULATE THE STABILITY OF COHESIN AND COHESIN-ANCHORED CONTACTS IN POST-MITOTIC ASTROCYTES

3.1. INTRODUCTION

Genome-wide Hi-C studies in human and mouse cells support the idea that cohesin has a widespread role in anchoring long-range chromatin interactions which in turn structure chromosomal domains (Seitan et al., 2013; Sofueva et al., 2013; Zuin et al., 2014). Despite these observations, it was nevertheless difficult to establish cohesin's causal role in facilitating domain structure due to its many direct and indirect effects. While cohesin-mediated interactions were significantly reduced in Rad21-deficient post-mitotic cells, they were not abolished after cohesin depletion. Several explanations are possible; including 1) additional proteins are required for the maintenance of domain structure, 2) domain structure is simply maintained in the absence of cohesin proteins or, 3) there is sufficient 'residual' cohesin in our post-mitotic system to maintain the observed chromatin interactions. I hypothesised that a population of residual cohesin complexes exist in Rad21-deficient post-mitotic cells and that these are sufficient to maintain the residual chromatin domain structures observed in these cells.

Residual cohesin could arise due to inefficient depletion experiments, or because a subset of protein is extremely long-lived. Indeed, Rad21 DNA and RNA levels were efficiently deleted (described in CHAPTER 2). Meanwhile, evidence from the literature supports the existence of stable pools of cohesin protein on chromatin. In mouse meiosis, cohesin can entrap sister chiasmata for years (Tachibana-Konwalski et al., 2010). In mitosis, Gerlich *et al.* used FRAP analysis to show evidence for two G_2 populations of cohesin in HeLa cells, a 'stable' pool with a residence time of 25 minutes and a more 'dynamic' pool with a residence time of 6 hours (Gerlich et al., 2006). In G_1-synchronised HeLa cells, Ladurner *et al.* showed an additional 'transient' pool of cohesin with a residence time of less than 1 minute in addition to the 'dynamic' pool reported previously (Ladurner et al., 2014) agreeing with what it was shown in *Drosophila* (Gause et al., 2010).

During the cell cycle, the regulation of cohesin association with chromatin involves the interplay of several regulators and modifications of the core cohesin complex. Cohesin is loaded onto chromatin by the loading factor in an ATP-dependent manner (Ciosk et

al., 2000; Ladurner et al., 2014). In G_1, cohesin 'dynamic' pools depend on a balance between the loading by Nipbl and unloading reaction by Wapl and Pds5 (Gandhi et al., 2006; Kueng et al., 2006). Important for cohesin's association with chromatin are Stag proteins (Murayama and Uhlmann, 2014; Remeseiro et al., 2012b; Sumara et al., 2000). During S-phase, sororin is recruited by the acetylation of Smc3 and binds to Pds5 counteracting the Wapl-unloading reaction (Nishiyama et al., 2010). This mechanism and the acetylation of Smc3 during DNA replication by Esco proteins (Beckouet et al., 2010; Ben-Shahar et al., 2008; Zhang et al., 2008) is proposed to lead to the previously discussed 'stable' pool of 'cohesed' cohesin. In *Drosophila* G_2 cells, it was shown by FRAP studies that changes in the levels of Pds5 and Nipbl leads to distinct changes in a proportion of 'stable' cohesin on chromatin (Gause et al., 2010). In vertebrates, these mechanisms are also essential for cohesin removal and SCC dissolution during mitosis (Waizenegger et al., 2000).

Thus, I hypothesised that de-stabilisation of the residual cohesin complexes off of chromatin would lead to distinct changes in chromosome organisation. Thus, in this CHAPTER, I explored the possibility that known cohesin regulators may have a role in sustaining residual complexes on chromatin.

3.2. RESULTS

RAD21-DEPLETED POST-MITOTIC ASTS HAVE RESIDUAL COHESIN ON CHROMATIN

Rad21-deficient cells had residual chromatin contacts arising from Rad21 binding sites (**Figure 28**). I hypothesised that these residual contacts were anchored from residual cohesin protein and investigated further their existence in Rad21-deficient ASTs.

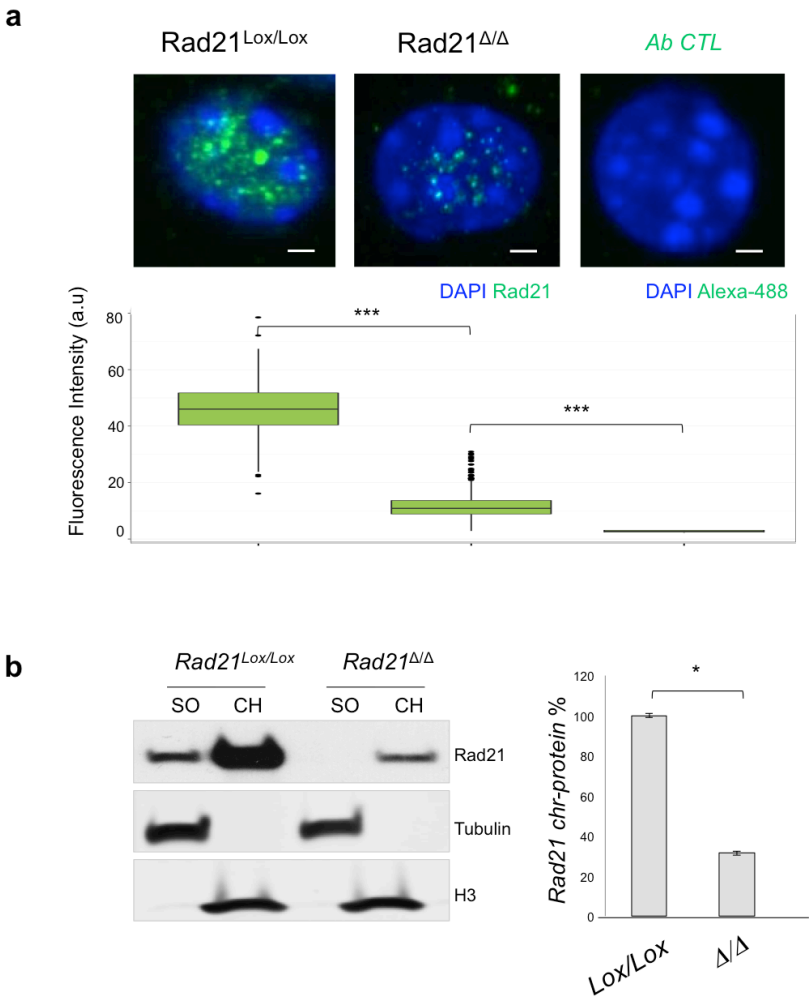

Figure 29. Cohesin complexes exist on chromatin in Rad21-deficient ASTs. a) Maximum projections of immunocytochemistry experiments and confocal imaging of AST *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ nuclei stained for Rad21 (green) after soluble proteins were extracted with 0.1% Triton-X100. DNA was counterstained with DAPI (blue) and Alexa-488 was used as a secondary antibody control. (Scale bar=2.5 µm). Whiskers and boxes indicate all and 50% of values respectively of Rad21 intensity measurements using the 'find object' feature of Volocity software and they were: 45 (*Lox/Lox*), 10 (*Δ/Δ*) in arbitrary units (N_{nuclei}=250 from 3 independent experiments). Statistical significance has been assessed by Wilcoxon Rank Sum and Signed Rank Test (Lox/Lox to *Δ/Δ* : $p<2.2x10^{-16}$; Ab CTR to *Δ/Δ* : $p<2.2x10^{-16}$). **b)** Western blot analysis for Rad21, tubulin and Histone 3 (H3) protein in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ ASTs. Samples were collected for protein analysis after fractionation into soluble (SOL) and chromatin-bound (CH) protein fractions. Bargraph represents the percentage of chromatin-bound Rad21 normalised to H3 levels in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ ASTs (33±3% from 3 independent experiments). Statistical significance has been assessed by Wilcoxon Rank Sum and Signed Rank Test (*Rad21*$^{Δ/Δ}$ 96h p=0.05).

To explore this, I performed immunofluorescence experiments for Rad21 protein after pre-extraction of soluble proteins. This approach allowed the quantification specifically of the remaining chromatin-bound Rad21 protein in *Rad21*$^{\Delta/\Delta}$ ASTs (**Figure 29**). This analysis showed that on average 25% of the total Rad21 chromatin-bound pool is residual in *Rad21*$^{\Delta/\Delta}$ ASTs (Nnuclei=250, from 3 independent experiments, $p<2.2x10^{-16}$ assessed by Wilcoxon Rank Sum and Signed Rank Test) (**Figure 29a**). This was confirmed by protein fractionation experiments, which were designed to follow the same principle, revealing that 33±3% of Rad21 protein is bound to chromatin in *Rad21*$^{\Delta/\Delta}$ ASTs (from 3 independent experiments, p=0.05 assessed by Wilcoxon Rank Sum and Signed Rank Test) (**Figure 29b**). Therefore, the existence of 'residual' cohesin complexes was confirmed in Rad21-deficient ASTs.

RESIDUAL COMPLEXES ARE SALT-RESISTANT AND ENRICHED FOR SPECIFIC COHESIN REGULATORS

As mentioned previously, there are several known regulators of cohesin stability on chromatin, primarily during S- and G_2-phase. These include the loading (i.e. Nipbl) (Ciosk et al., 2000; Ladurner et al., 2014) and the anti-establishment factors (i.e. Wapl) (Gandhi et al., 2006; Kueng et al., 2006) or cohesin modifications such as the acetylation of Smc3 by Esco is known to stabilise cohesin on chromatin by creating a cohesive form during S-phase and 'locking' sister chromatids together (Beckouet et al., 2010; Unal et al., 2008). In G_1, the function of these regulators in cohesin functions in gene regulation is still largely unknown. An additional interphase regulator of cohesin on chromatin is CTCF (Parelho et al., 2008; Schmidt et al., 2010) and cohesin complexes interact with the C-terminus of CTCF through the Stag proteins (Xiao et al., 2011). I explored whether specific cohesin regulatory factors may be enriched in the chromatin-bound pool in *Rad21*$^{\Delta/\Delta}$ ASTs and thus contributing to their stability on chromatin.

I first evaluated the expression levels of the main components known to regulate cohesin chromatin binding by taking advantage of the RNA-Seq data generated previously (**Figure 17-20**). The expression of cohesin core components and regulators was reduced in ASTs compared to NSCs (**Figure 30a**). The mRNA levels of Rad21, Smc1a, Smc3, Stag2, Pds5a/b, Sororin were reduced by up to 50% in ASTs, and the expression of two factors, Esco2 and separase was undetectable whereas Esco1 was still expressed. Both Esco1 and Esco2 are reported to be responsible for cohesin

stabilisation on chromatin during S-phase (Hou and Zou, 2005; Nishiyama et al., 2010) and the protease separase removes centromere cohesin by mediating its proteolytic cleavage of Rad21 at the onset of anaphase (Nakajima et al., 2007; Uhlmann et al., 1999). Given these proteins are responsible for SCC, the expression of Esco1 in our post-mitotic cells was very unexpected. Their mRNA expression remained unchanged upon *Rad21* genetic depletion in both cell types and only Rad21 levels were reduced to 7%. Thus, both cycling and post-mitotic Rad21-deficient cells expressed known regulators of cohesin association with chromatin.

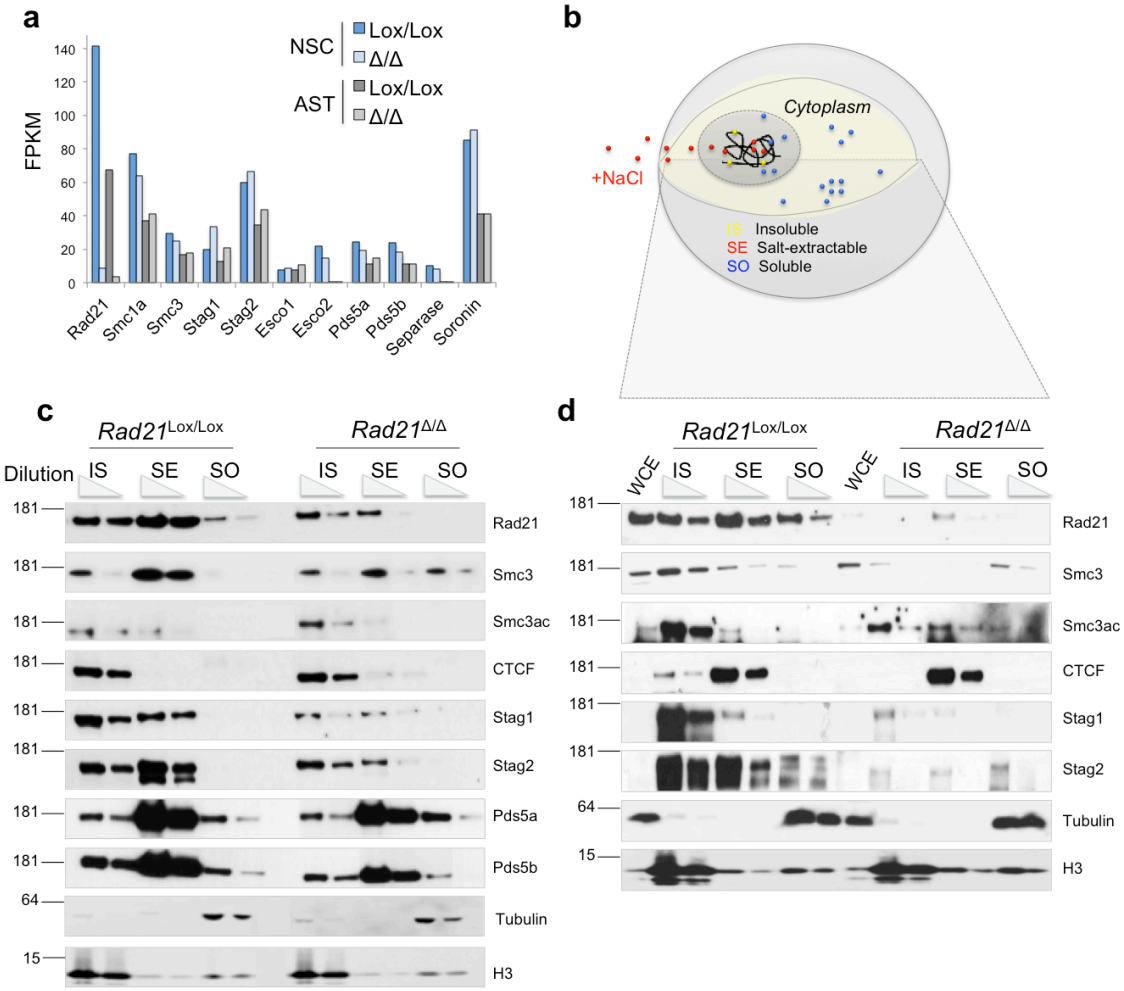

Figure 30. Cohesin regulators are expressed and present with residual cohesin complexes in Rad21-deficient ASTs. a) mRNA expression levels of cohesin proteins expressed as FPKM values in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ NSCs and ASTs (in blue and grey). **b)** Graphical representation of the standard protein fractionation method enriched with a salt-extraction step. This method isolates the cytoplasmic (SOL, blue) from the chromatin-bound protein fraction, and which is then separated into a salt-extractable (SE. red) and insoluble (IS, yellow) (Martini et al., 1998). **c,d)** Western blot analysis for Rad21, Smc3, Smc3 LysAc, CTCF, Stag1/2, Pds5a/b, tubulin, and histone 3 (H3) protein levels in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ ASTs. Samples were collected for protein analysis after soluble protein fractions were separated from chromatin fractions and a salt-extracted chromatin fraction was prepared by the addition of 300 and 600 mM NaCl, respectively. WCE: Whole cell extract

To further characterise the residual chromatin-bound complexes at the protein level, a standard protein fractionation technique enriched with a salt-extraction step was used (**Figure 30b**). This method allowed for the isolation of chromatin bound proteins into salt-extractable and insoluble protein fractions (Martini et al., 1998). We reasoned that insoluble proteins are more stably bound to chromatin than proteins, which can be extracted with high salt. As controls, Histone 3 (H3) was enriched in the insoluble fraction and tubulin was enriched in the soluble fraction. Initially, the experiment was performed with 300 mM salt to the chromatin bound fraction of both *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ ASTs (**Figure 30c,d**). As before, Rad21 protein was reduced to 80% in *Rad21*$^{\Delta/\Delta}$ ASTs (**Figure 14c**). In *Rad21*$^{Lox/Lox}$ ASTs, Rad21 was primarily enriched in the salt-extractable fraction, whereas in *Rad21*$^{\Delta/\Delta}$ cells, the remaining Rad21 was enriched in the insoluble fraction. As expected, another cohesin core component, Smc3, was shown to translocate from the chromatin-bound fraction to the soluble fraction in *Rad21*$^{\Delta/\Delta}$ ASTs (Laugsch et al., 2013). CTCF was found in the insoluble fraction in both *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ ASTs and the total amounts of CTCF remained unchanged in both conditions (**Figure 30c**). Stag1 was found mainly in the insoluble fraction and after Rad21 depletion, it remained enriched in the same fraction, however its total levels were significantly decreased. Stag2 showed enrichment in the salt-extractable fraction and after Rad21 depletion, it was enriched at the insoluble fraction while its total levels were decreased similarly to Stag1. Pds5a and b were both enriched at the salt-extractable fraction but their total levels were not affected significantly by Rad21 depletion. Acetylated-Smc3 was enriched in the insoluble fraction and remained enriched here in the Rad21-deficient ASTs. These results showed that the remaining cohesin complexes in *Rad21*$^{\Delta/\Delta}$ ASTs were primarily found in the insoluble fraction of chromatin, were enriched for marks associated with stable cohesin and were associated with an insoluble CTCF pool. When the fractionations were performed in 600 mM salt, the insoluble chromatin-bound Rad21 fraction in *Rad21*$^{\Delta/\Delta}$ ASTs was enriched for Stag1, Smc3ac and CTCF, further supporting the above results (**Figure 30d**).

Thus, these results support the initial hypothesis, which predicts that there are residual chromatin-bound cohesin complexes and further implied that this residual pool may represent a stable chromatin fraction.

CTCF AND STAG DEPLETION DESTABILISES RESIDUAL CHROMATIN-BOUND COHESIN IN RAD21-DEFICIENT ASTS

CTCF and Stag proteins are enriched in the chromatin fraction along with residual Rad21 in *Rad21*$^{\Delta/\Delta}$ ASTs. Thus, I hypothesised that depletion of CTCF or Stag proteins may lead to the destabilisation of residual cohesin complexes on chromatin in *Rad21*$^{\Delta/\Delta}$ ASTs.

To knock-down CTCF and Stag1/2 in *Rad21*$^{\Delta/\Delta}$ ASTs, siRNA oligos against CTCF or Stag mRNA were transfected into *Rad21*$^{\Delta/\Delta}$ ASTs 72 hours post-infection (**Figure 31a**). The efficiency of transfection in both *Rad21*$^{\Delta/\Delta}$ and *Rad21*$^{Lox/Lox}$ ASTs was evaluated using siGLO-488 (siRNA-attached to a fluorochrome, described in CHAPTER 5) reagent. Using flow cytometry as a read-out, more than 90% of ASTs were successfully transfected (**Figure 31b**).

To determine the optimal time of knockdown using siRNAs in ASTs, I transfected siRNAs and collected cells at various timepoints post-transfection to evaluate protein levels. Specifically, Stag1 and Stag2 were considerably depleted from chromatin by 2 days post-transfection (**Figure 32a,b**). The same timepoint was chosen for CTCF (**Figure 32c**). Stag1 knockdown did not affect the levels of Stag2, confirming the specificity of the knockdowns, especially as these are highly conserved proteins (Roig et al., 2014). Previous studies have shown that the depletion of Stag1 doesn't affect Stag2 levels (Xiao et al. 2011, Remeseiro et al., 2012).

Of the residual Rad21, a further 91%, 87% and 84% was depleted after *Rad21*$^{\Delta/\Delta}$ ASTs were treated with CTCF, Stag1 and Stag2 RNAi for 3 days, respectively (**Figure 32b**) (*Rad21*$^{\Delta/\Delta}$ to siCTCF, siStag1, siStag2: 91±2.8%, 89±6.8%, 84±6.7%, from 3 independent experiments, respectively, p=0.05 assessed by Wilcoxon Rank Sum and Signed Rank Test). Due to experimental variability, statistical significance between control and siRNA-treated *Rad21*$^{\Delta/\Delta}$ was not established but this analysis already suggests that the depletion of CTCF and Stag1/2 in *Rad21*$^{\Delta/\Delta}$ ASTs leads to the destabilisation of the residual Rad21.

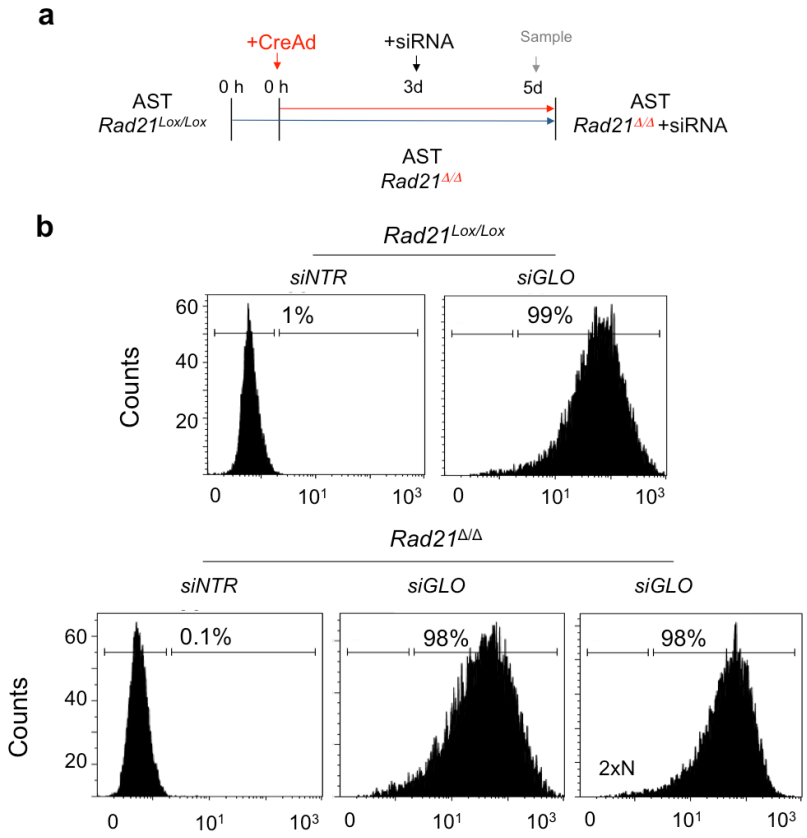

Figure 31. Post-mitotic Rad21-deficient ASTs are successfully transfected with siRNA oligos against cohesin regulators. a) Graphical representation of the protocol used to knock-down cohesin regulators in *Rad21*$^{Lox/Lox}$ ASTs. ASTs were infected with Cre-expressing adenoviruses (Cre-Ad) and 72 hours post-infection, they were transfected with siRNAs against the cohesin regulator mRNA. Grey arrowheads represent timepoint when samples were collected. **b)** Transfection efficiency by flow cytometry for green fluorescence after 48 hours of transfection with either siRNA non-targeting control (siNTR) or siGLO-488 (siRNA-attached to a fluorochrome) (50 nM) in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$ ASTs. ASTs were plated at densities of 0.4x10^{6} per 6-well and at a two-fold density in *Rad21*$^{Δ/Δ}$ ASTs (2xN) was tested.

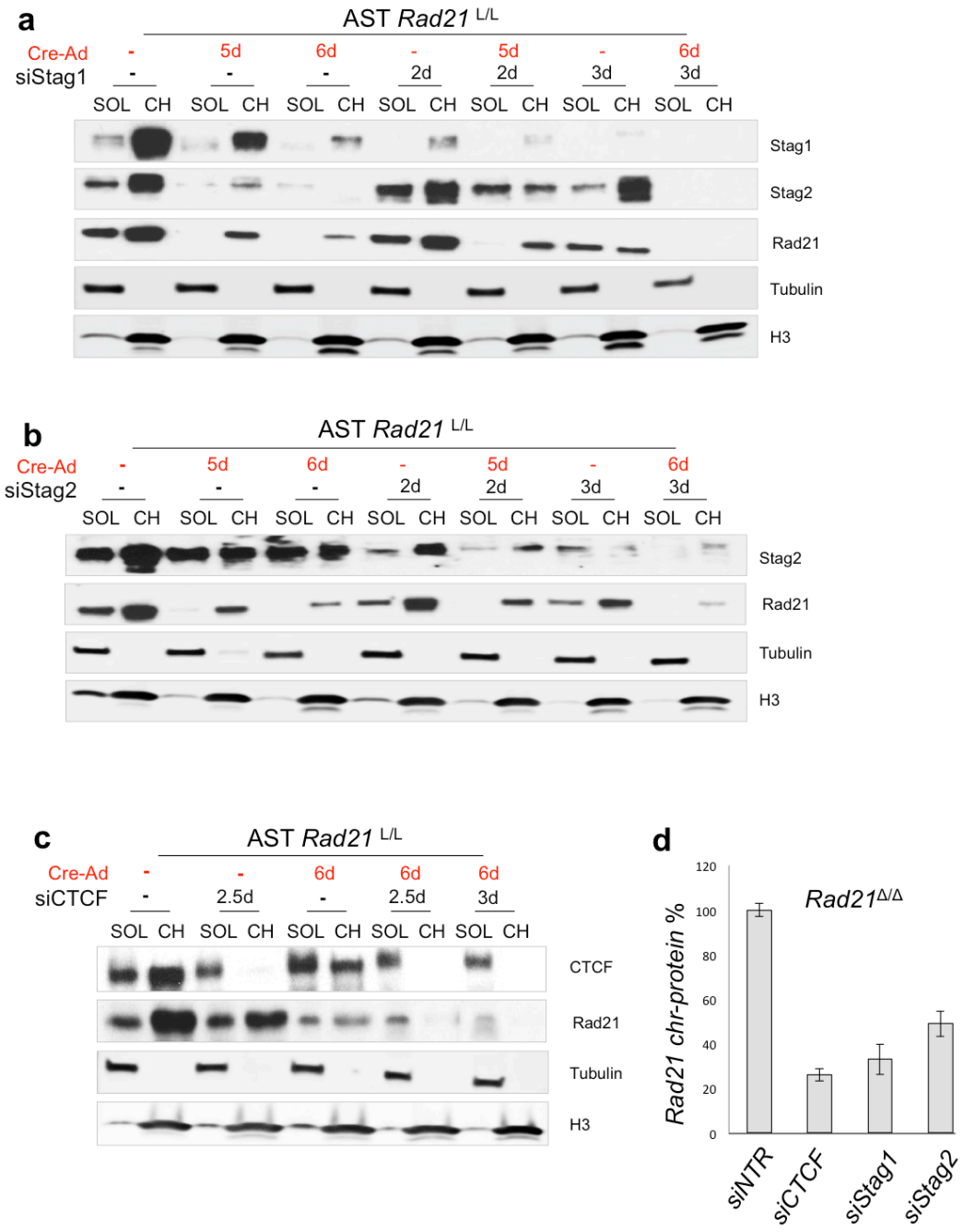

Figure 32. Depletion of cohesin regulators results in cohesin destabilisation off chromatin in control and Rad21-deficient ASTs. a) Western blot analysis for Stag1, Stag2, Rad21, H3 and tubulin protein in a timecourse of siRNA for Stag1 transfection in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ ASTs. **b)** Western blot analysis for Stag2, Rad21, H3 and tubulin protein in a timecourse of siRNA for Stag2 transfection in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ ASTs. **c)** Western blot analysis for CTCF, Rad21, H3 and tubulin protein in a timecourse of siRNA for CTCF transfection in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ ASTs. All samples were collected for protein analysis after fractionation into soluble and chromatin-bound protein fractions (SOL: Soluble proteins, CH: Chromatin-bound proteins). **d)** Graph represents the percentage of chromatin-bound Rad21 normalised to H3 levels for siNTR and siCTCF, siStag1, siStag2 *Rad21*$^{\Delta/\Delta}$ ASTs (91±2.8%, 89±6.8%, 84±6.7%, from 3 independent experiments, respectively). Statistical significance has been assessed using a Wilcoxon Rank Sum and Signed Rank Test (p=0.05).

These experiments were repeated such that all knockdowns were performed from the same parental pool of *Rad21*$^{\Delta/\Delta}$ ASTs and protein levels could be assessed from all samples simultaneously (**Figure 33**). In addition to using siRNAs for CTCF, Stag1 and Stag2 as before, siRNAs to Esco1 were also included because the insoluble pool of residual Rad21 was enriched for Smc3 acetylation (**Figure 30c-d**). Depletion of CTCF caused the most pronounced decrease in Rad21 protein levels, with 91% of residual cohesin being depleted and a decrease in Stag1, Stag2 and Smc3 acetylation levels were also observed. These findings indicated that there were Smc3 acetylated residual cohesin-Stag1 or Stag2 complexes which were sensitive to CTCF depletion. As expected, Esco1 depletion resulted in undetectable levels of Smc3 acetylation. Interestingly, the depletion of Esco1 led to a 61% decrease in Rad21 levels compared to *Rad21*$^{\Delta/\Delta}$ ASTs and in both Stag1 and Stag2 levels without affecting CTCF levels on chromatin. This implied that both residual cohesin-Stag1 or Stag2 complexes were sensitive to the depletion of Smc3 acetylation. Co-immunoprecipitation experiments were attempted in order to confirm the composition of the residual complexes, however these were not successful due to the prohibitive amount of material required. Despite this, these results showed that Smc3 acetylated residual cohesin-Stag1 or Stag2 complexes were sensitive to CTCF and Esco1 depletion. Overall, they underlined the role of CTCF in maintaining their stability on chromatin.

Thus, I confirmed that the RNAi-mediated depletion of a series of cohesin regulators in cohesin-deficient ASTs led to further destabilisation of the residual chromatin-bound cohesin complexes. This analysis also revealed different sensitivities to the different regulators, highlighting that the residual complexes are most sensitive to CTCF depletion and uncovered a potential dependency of a fraction of them on Smc3 acetylation via the depletion of Esco1.

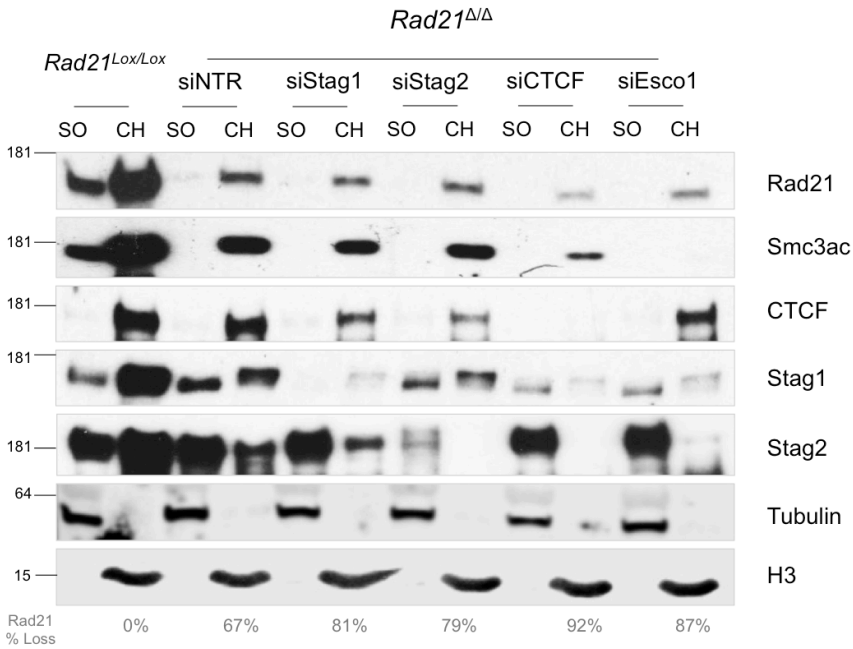

Figure 33. Residual chromatin-bound Rad21 is most sensitive to CTCF and Esco1 depletion in Rad21-deficient ASTs. Western blot analysis for Stag1, Stag2, Rad21, H3 and tubulin protein in *Rad21*$^{Lox/Lox}$ and transfected with siRNA for Stag1, Stag2, CTCF and Esco1 *Rad21*$^{\Delta/\Delta}$ ASTs for 3 days. Samples were collected for protein analysis after fractionation into soluble and chromatin-bound protein fractions (SOL: Soluble proteins, CH: Chromatin-bound proteins). Chromatin-bound Rad21 was normalised to H3 levels and the percentages of Rad21 loss calculated for the different samples in this experiment is indicated in grey (1 experiment).

RAD21 BINDING IS SENSITIVE TO CTCF DEPLETION IN COHESIN-DEFICIENT ASTS

To understand the effect of CTCF and Stag1 depletion on residual cohesin binding sites, I performed a genome-wide analysis of cohesin binding in CTCF-depleted *Rad21*$^{\Delta/\Delta}$ and Stag1-depleted *Rad21*$^{\Delta/\Delta}$ ASTs.

For this purpose, Rad21 ChIP-Seq libraries were prepared from control *Rad21*$^{Lox/Lox}$, *Rad21*$^{\Delta/\Delta}$, CTCF-depleted *Rad21*$^{\Delta/\Delta}$ and Stag1-depleted *Rad21*$^{\Delta/\Delta}$ ASTs. The ChIP-Seq protocol was further optimised so that the ChIP efficiency was increased and therefore allowed me to produce libraries from cells that have low amounts of Rad21 protein on chromatin (**Figure 32-33**). Optimisations included increasing the time of formaldehyde fixation from 10 (Sofueva et al., 2013) to 30 minutes and using three timepoints of sonication to test ChIP efficiency by western blot and ChIP-qPCR. The comparison between the different conditions and with known NSC conditions showed that 30 minutes fixation and 8 minutes sonication were optimal to achieve maximum efficiency of Rad21 ChIP in ASTs (**Figure 34b,d**). Next, I prepared ChIP-Seq libraries by using

equal amounts of DNA from the series of samples in the NEB Illumina ChIP-Seq Library preparation kit (NEB). The consequent sequencing in the Illumina MiSeq platform yielded comparable numbers and quality of sequencing single-end 75 bp reads for the different libraries (**Table 4** in the APPENDIX, 2 replicates per library).

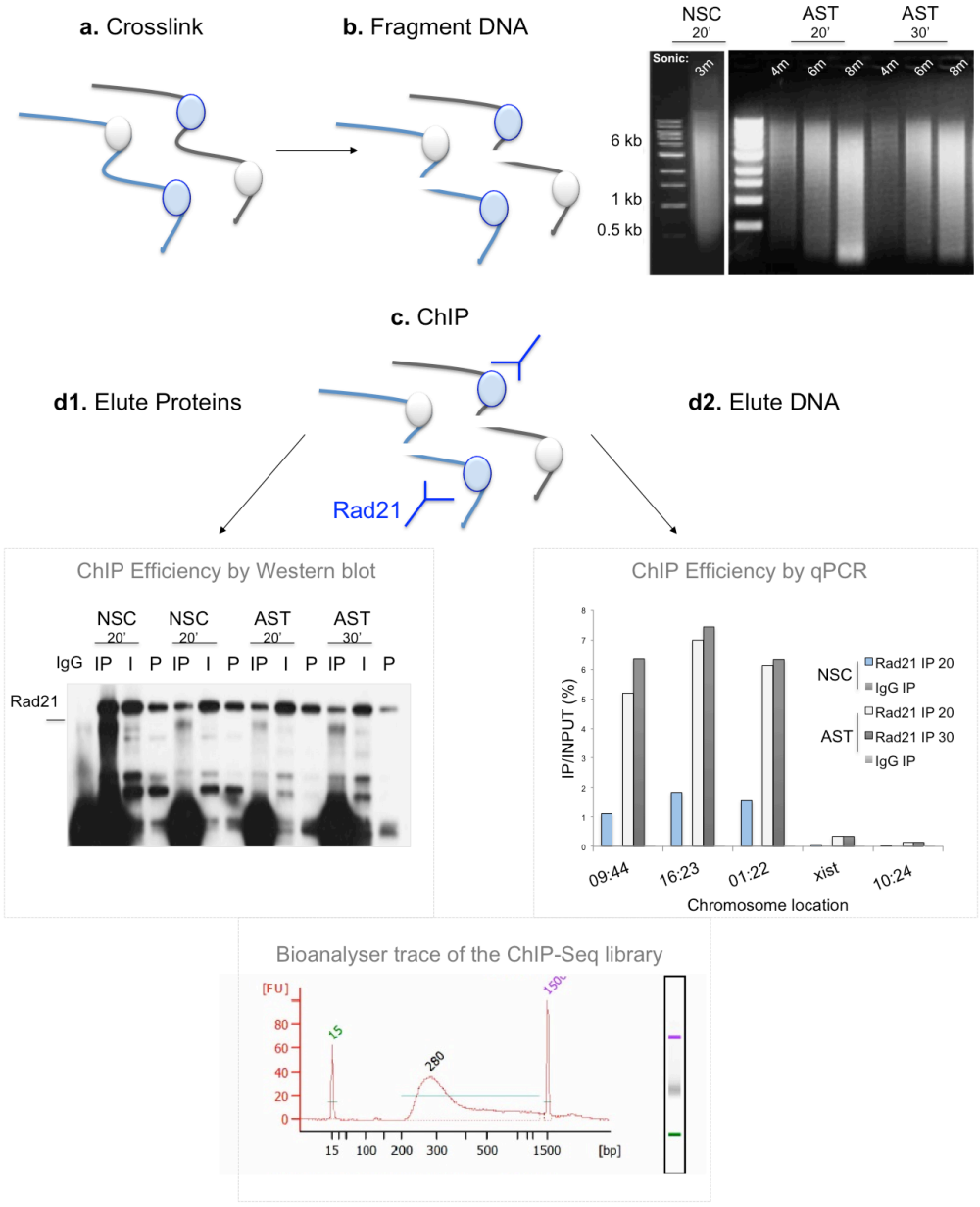

Figure 34.The ChIP-Seq protocol and quality control steps used to increase the efficiency of Rad21 chromatin-immunoprecipitation in NSCs and ASTs. Main steps of the protocol are as follows: **a)** crosslinking proteins with DNA using 1% folmadahyde (20,30 minutes); **b)** Chromatin fragmentation by sonication (4,6,8 minutes); **c)** Chromatin-Immunoprecipitation (ChIP) of Rad21 protein by using monoclonal Rad21 antibody (Ab922); **d1, d2)** Protein or DNA elution to test efficiency by western blot or ChIP-qPCR (09:44, 16:23, 01:22 positive and Xist, 10:14 negative control Rad21 sites). Finally, DNA purification, library preparation (NEB) (optimal size of the library is 300 bp) and sequencing via the Illumina standard protocol.

The analysis of the ChIP-Seq dataset was performed as described in (Sofueva et al., 2013) by Dr C. Barrington (a bioinformatician in the Hadjur lab), where mapped reads were assigned to 50 bp bins genome-wide which allows the unbiased investigation of the ChIP signal from tens of thousands of sites across the genome. The reads were transformed into percentiles log10(1-quantile(v)) to allow us to compare ChIP-seq libraries between datasets. ChIP libraries were well replicated for all datasets (**Figure 35a**), allowing us to choose robust thresholds for Rad21 peaks (i.e top 1% of ChIP-Seq signal). To determine which Rad21 peaks changed during treatment, signals in control *Rad21*$^{Lox/Lox}$ ASTs were plotted against each experimental condition (**Figure 35b**). This analysis revealed global changes in Rad21 signals in each experimental sample, with the largest changes in Rad21 signal observed in CTCF-depleted *Rad21*$^{\Delta/\Delta}$ ASTs.

To more carefully assess the impact of the knockdowns on different kinds of cohesin binding sites, Rad21 sites were grouped according to signal strength in *Rad21*$^{Lox/Lox}$ ASTs into high, medium and low signal sites. For this purpose, the top, medium, and lower 10% of signal values were utilised. The distribution of the ChIP signal in each group is shown for each of the experimental samples (**Figure 35c**). As expected, the median signal for *Rad21*$^{\Delta/\Delta}$ ASTs was less than in control *Rad21*$^{Lox/Lox}$ cells, evident at each signal group (**Figure 35c,d**). Thus, a decrease in overall protein levels (as seen by Western blots in **Figure 32-33**) was accompanied by decreased Rad21 ChIP signals. Stag1-depleted *Rad21*$^{\Delta/\Delta}$ ASTs had an effect on Rad21 signal which was similar to the effect observed in *Rad21*$^{\Delta/\Delta}$ ASTs. However, CTCF-depleted *Rad21*$^{Lox/Lox}$ ASTs had the most pronounced effect on Rad21 signal in all signal categories compared to *Rad21*$^{Lox/Lox}$ ASTs. Similarly, looking at an example region in the genome browser, the ChIP-Seq data again show that CTCF knock-down in *Rad21*$^{\Delta/\Delta}$ ASTs has the greatest effect on cohesin binding (**Figure 35d**).

Hence, a genome-wide analysis of cohesin binding sites revealed that a further decrease in cohesin protein levels translated into a decrease in cohesin ChIP signal upon the depletion of cohesin regulator proteins. This analysis emphasised a role for CTCF in all ChIP signal ranges, even in the high-signal Rad21 sites, which are more likely to regulate domain structure.

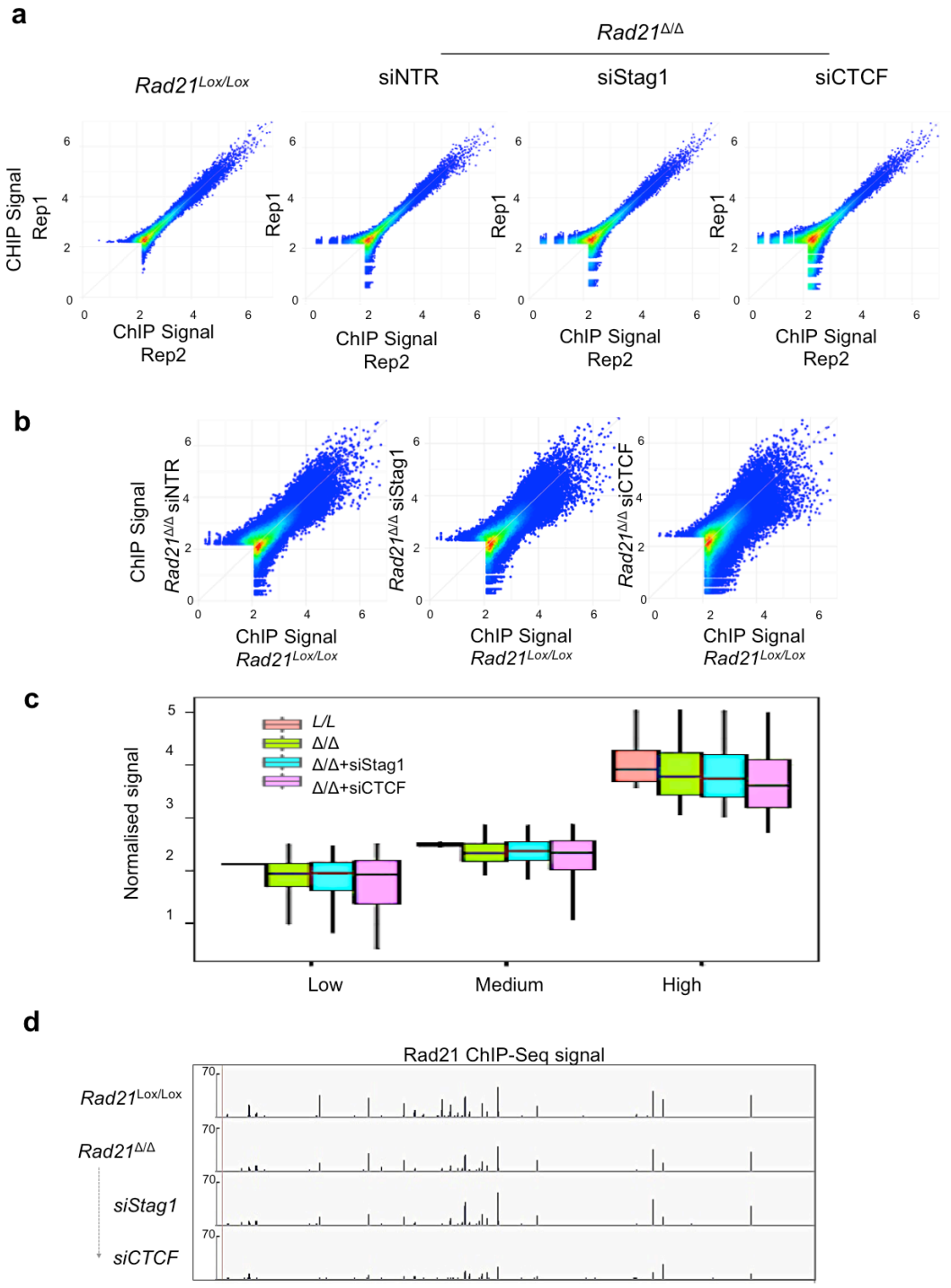

Figure 35. Rad21 binding profiles are most sensitive to CTCF depletion in Rad21-deficient ASTs. a) Correlation of Rad21 ChIP signal (50 bp) after quantile normalisation between the two replicates of control *Rad21*$^{Lox/Lox}$ and *Rad21*$^{Δ/Δ}$, CTCF- and Stag1-depleted *Rad21*$^{Δ/Δ}$ ASTs (top 1% of bins) and **b)** between *Rad21*$^{Lox/Lox}$ and control, CTCF-, Stag1-depleted *Rad21*$^{Δ/Δ}$ ASTs, respectively (top 1% of bins). **c)** Using the top (High), medium (Medium) and lower (Low) 10% of signal bins in *Rad21*$^{Lox/Lox}$ ASTs, the ChIP signal in the different samples (*Rad21*$^{Lox/Lox}$ and CTCF-, Stag1-depleted, control *Rad21*$^{Δ/Δ}$ ASTs) were plotted. **d)** Representation of ChIP-Seq reads aligned in mouse genome (mm10) as they are represented in the Genome Browser in *Rad21*$^{Lox/Lox}$ and control, CTCF-, Stag1-depleted *Rad21*$^{Δ/Δ}$ ASTs.

STAG1 AND CTCF DEPLETION IN COHESIN-DEFICIENT ASTS COMPROMISES CHROMATIN DOMAIN STRUCTURE

Previous results indicated that CTCF/cohesin ChIP signal correlates well with the extent of contact insulation (Sofueva et al., 2013), in other words, strong ChIP signals are associated with strong 'domain borders'. Given that CTCF-depleted *Rad21*$^{\Delta/\Delta}$ ASTs had a pronounced effect on Rad21 signal, including in the 'high signal' category, I explored the effect of CTCF, Stag1 and Stag2 loss on residual chromatin structure in *Rad21*$^{\Delta/\Delta}$ ASTs.

To do this, I employed the *in nucleo* Hi-C method previously used in the lab (Sofueva et al., 2013). As before, the Hi-C method is an unbiased analysis of global chromosome organisation and is based on the same 'proximity-ligation' principle used for all 'C' techniques (Dekker et al., 2002) (**Figure 36**). Hi-C is different from 4C methods in that ligation junctions are labeled with biotinylated nucleotides, which allows for the enrichment of chromatin fragments which were thought to be previously proximal in nuclear space (Lieberman-Aiden et al., 2009). While Hi-C libraries were prepared form the siStag2 and siNon-targeting control (siNTR) *Rad21*$^{\Delta/\Delta}$ ASTs with no changes to the standard protocol, the siCTCF and siStag1 cells required significant optimisations. The CTCF- and Stag1-depleted *Rad21*$^{\Delta/\Delta}$ samples were over-digested and the consequent ligation efficiency was poor (**Figure 36a-c**). After the decrease in digestion time to 1 day and a 2-fold increase in T4 DNA Ligase concentration, the ligation efficiency was similar between all libraries. Next, Hi-C libraries prepared from *Rad21*$^{Lox/Lox}$, *Rad21*$^{\Delta/\Delta}$, CTCF- depleted *Rad21*$^{\Delta/\Delta}$, Stag1-depleted *Rad21*$^{\Delta/\Delta}$ and Stag2-depleted *Rad21*$^{\Delta/\Delta}$ ASTs were sequenced using the Illumina HiSeq 2500 platform. Comparable number of paired-end 50 bp reads were obtained between the datasets. After alignment and filtering for biases (performed by Dr C. Barrington) the final Hi-C contact maps were prepared from similar numbers of read pairs (Library information in **Table 4** in the APPENDIX).

Results from the ChIP-seq analysis (**Figure 35c**) revealed that high signal Rad21 binding sites may be affected in CTCF and Stag1-depleted *Rad21*$^{\Delta/\Delta}$ ASTs. Furthermore, high-signal Rad21 sites were associated with strong contact insulation, or 'domain borders' (Sofueva et al., 2013) as these were defined in CHAPTER 1. Thus, we performed a global *contact insulation* analysis from the perspective of cohesin/CTCF binding sites in the genome after depletion for CTCF, Stag1 or Stag2. Contact insulation is a computational method, which allows the quantification of the average frequency of contacts between all loci that lie within a certain distance 'band'

(here 80-120 kb) on either side of a Rad21/CTCF binding site ('0' point). When the distribution of contact frequency is plotted, insulation at Rad21 sites is visualised by fewer contacts crossing the '0' point compared to the neighboring regions (**Figure 37a**).

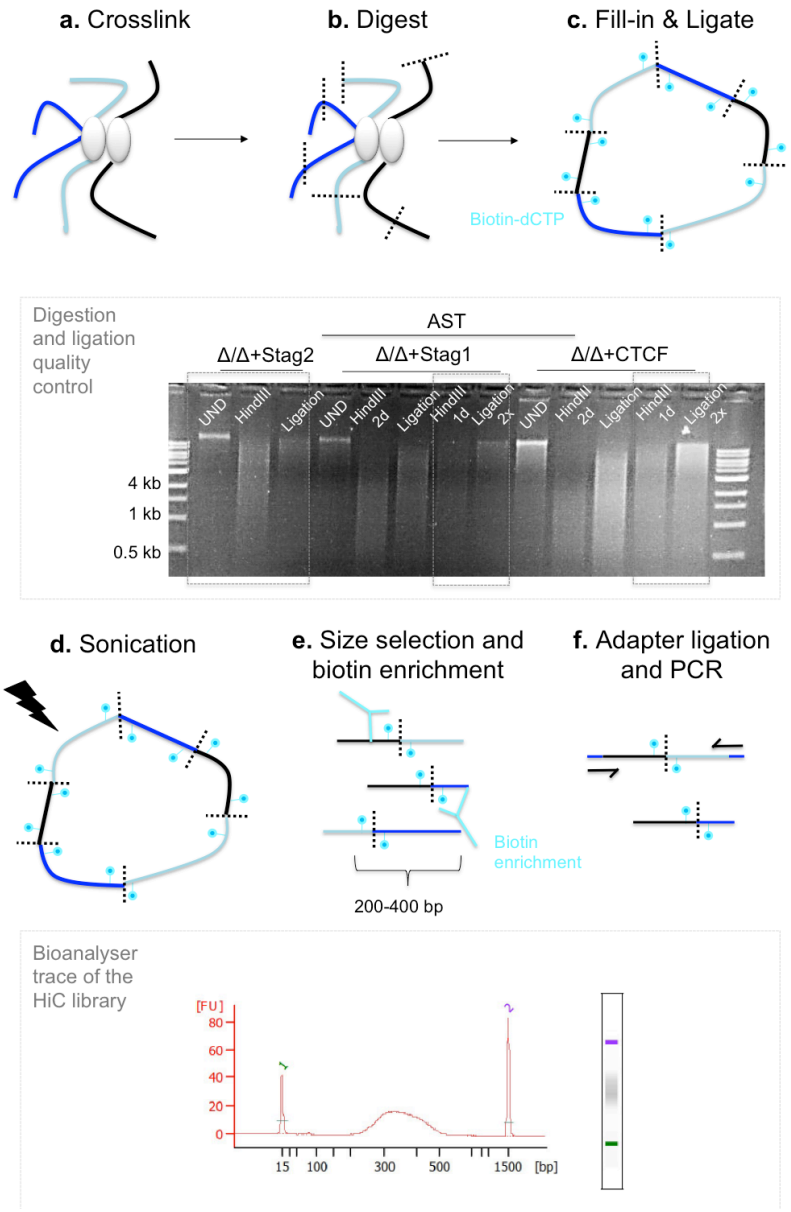

Figure 36. The HiC protocol and quality control steps used to optimise efficiency in siRNA-treated ASTs. Main steps of the protocol were as follows: **a)** crosslinking proteins with DNA using 1% formadahyde (20, 30 minutes); **b,c)** Digestion of DNA fragments with HindIII, fill-in of the nicked ends with biotin-dCTP and ligation reactions (1x and 2x Ligase concentration); **d,e)** Sonication of the ligation products, size selection (200-400 bp) and enrichment of the biotin-containing fragments; f) Illumina paired-end adapter ligation and PCR reactions (300-500 bp) followed by sequencing. UND: Undigested sample, HindIII 1d and 2d: Digested sample with HindIII for 1 and 2 days, Ligation 2x: Ligated samples with double the amount of T4 Ligase.

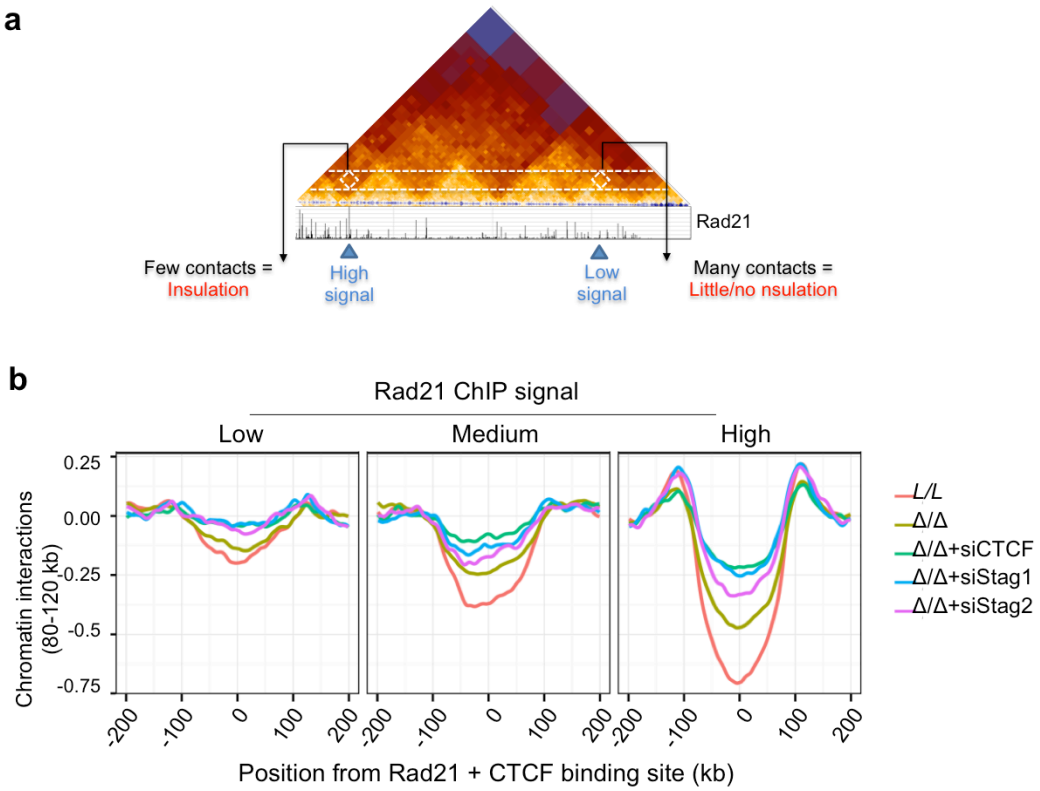

Figure 37. Stag2, Stag1 and CTCF depletion leads to a progressive loss in contact insulation in Rad21-deficient ASTs. a) A schematic representation of the method used to calculate *contact insulation* for a given locus in the genome. If insulation is occurring from a given point in the genome (green arrows), then you detect fewer contacts crossing the point compared to the neighboring regions. This depletion of contacts represents *contact insulation*. Conversely, if many contacts are detected across a point this represents little or no insulation (compare white triangles in Hi-C matrix, top). Using the Hi-C matrix, the average intensity of contacts is calculated between all chromosomal elements that lie within a certain distance band (white horizontal lines) on either side (e.g. +/-200 kb) of a given feature in genome, for example Rad21 binding sites (green arrows). This distribution of contact frequency is plotted below the Hi-C matrix. **b)** specifically 400 bp region centered about Rad21 binding sites (points 0) and stratified by Rad21 signal in the *Rad21*$^{Lox/Lox}$ ASTs ChIP-Seq dataset to 10% of the Low, Medium and High signals (from left to right panels), co-localising with CTCF sites (ChIP-signal>2.3).

In this analysis, insulation from Rad21 co-localising with CTCF sites (mouse cortex; ENCODE) was calculated for the different Rad21 signal intensity categories, namely 'Low', 'Medium' and 'High' which were described previously (**Figure 35c**). Rad21 binding sites with 'Low' and 'Medium' signal were shown to have minimal genome-wide insulation profile but the 'High' signal sites mediate strong insulation confirming what was shown before (Sofueva et al., 2013) (**Figure 37b**). Interestingly, contact insulation progressively decreased from control *Rad21*$^{Lox/Lox}$ to control *Rad21*$^{\Delta/\Delta}$, Stag2-, Stag1- and CTCF-depleted *Rad21*$^{\Delta/\Delta}$ (**Figure 37b**). This analysis uncovered a significant drop in cohesin-mediated insulation across the genome upon the depletion of cohesin regulators while pointed to a different sensitivity to the different factors.

As an example for the global effects described above, we selected sites in the genome based on the change of Rad21 signal after knockdown with CTCF, Stag1 or Stag2. In the first example (**Figure 38**), a Hi-C contact map is shown for a 4 Mb region on chromosome 1. In addition are shown the CTCF and Rad21 ChIP-Seq signals. In control *Rad21*$^{Lox/Lox}$ ASTs, this region had several chromatin domains at the megabase and sub-megabase scales flanked by high signal cohesin and CTCF sites. As observed, Hi-C maps prepared from *Rad21*$^{\Delta/\Delta}$ ASTs had a loss of contacts within domains without a significant change to the chromosome domain borders. Knockdown of Stag2 in *Rad21*$^{\Delta/\Delta}$ ASTs did not dramatically change the domain structure compared to *Rad21*$^{\Delta/\Delta}$ ASTs. Meanwhile, depletion of CTCF and Stag1 in *Rad21*$^{\Delta/\Delta}$ ASTs leads to a dramatic change in domain structure in this region. Importantly this loss of domain structure was associated with Rad21 binding sites which were themselves lost, revealing that the loss of specific Rad21 binding sites leads to a corresponding loss of long-range contacts and domain structure.

In another example, I examined the Pcdh gene cluster, where I had previously performed 4C-seq analysis and could confirm the presence of cohesin-anchored enhancer-promoter chromatin loops (**Figure 39**). In **Figure 39**, a wider region around the Pcdh cluster is flanked by several cohesin and CTCF binding sites which were shown to be disrupted in the Pcdhα cluster and which in turn disrupted its regulation and choice of promoters (described in CHAPTER 2). The additional depletion of CTCF and Stag1 resulted in ectopic chromatin contacts from the Pcdhβ cluster to the downstream domain resulting from a widespread loss of cohesin sites spanning both the cluster and the downstream domain. Therefore, the Hi-C results showed that the depletion of CTCF and Stag1 in *Rad21*$^{\Delta/\Delta}$ ASTs led to the loss of Rad21 binding at specific sites in the genome and thus was accompanied with a corresponding loss in the chromatin contacts from these cohesin binding sites.

Therefore, the depletion of cohesin regulators in Rad21-deficient ASTs resulted in loss of cohesin binding sites, which was correlated in a genome-wide analysis with striking decrease in cohesin-mediated contact insulation. The results were supportive of an important role for CTCF in all signal strengths.

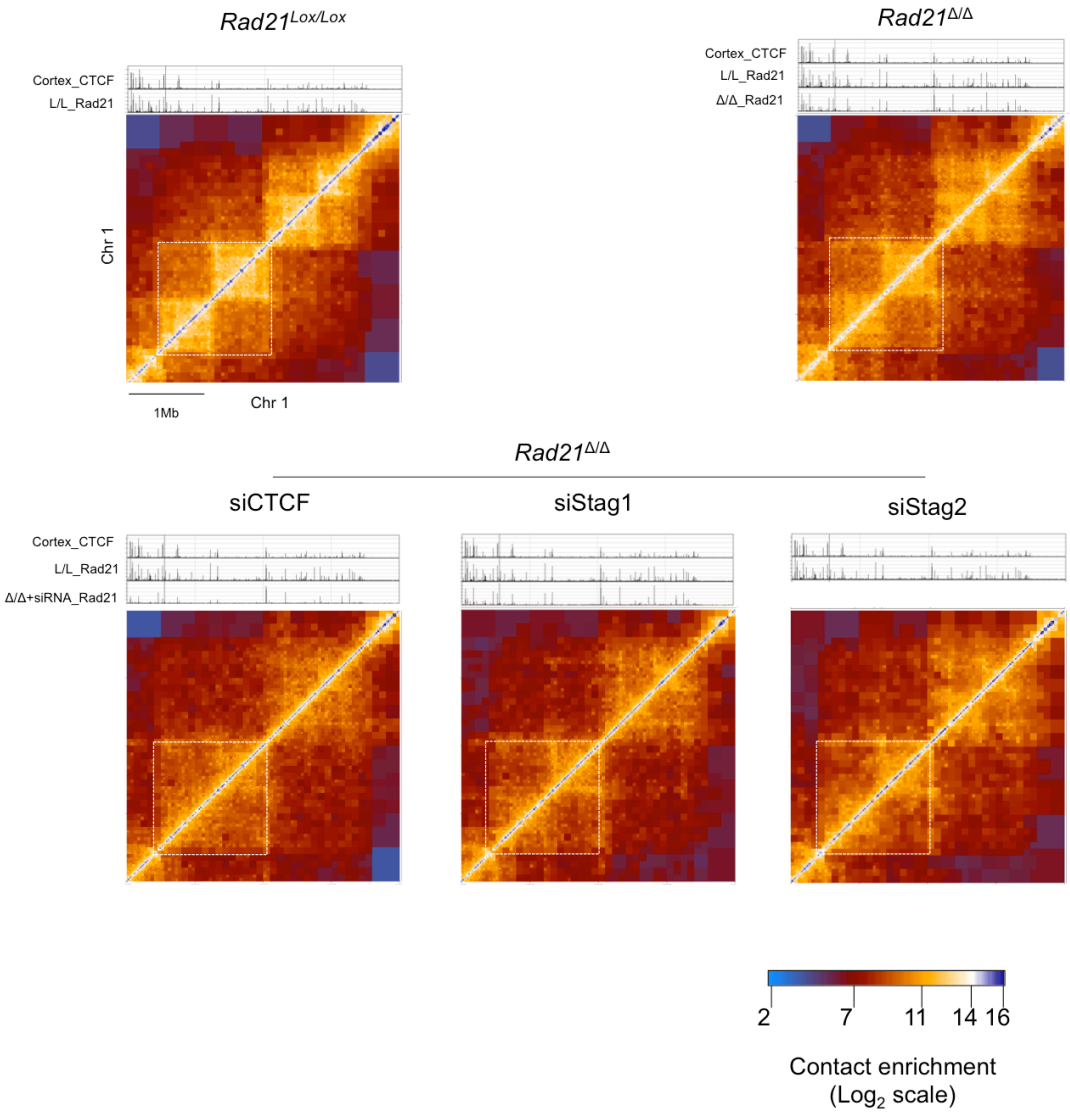

Figure 38. CTCF and Stag depletion lead to the dissolution of sub-domains, internal to a 4 Mb domain, accompanied by the loss of cohesin binding in cohesin-deficient ASTs. H-iC contact maps of a wide 4 Mb region on chromosome 1, where two sub-domains are located within each of the two 1.5 Mb domain flanked all by Rad21 and CTCF binding sites in *Rad21*$^{Lox/Lox}$ and in CTCF-, Stag1, Stag2-depleted, control *Rad21*$^{Δ/Δ}$ ASTs. Aligned are the genes spanning the region (top panel in Rad21$^{Lox/Lox}$), CTCF and Rad21 ChIP-Seq signal (top 1%) in the samples indicated on the left side of the tracks. CTCF ChIP-Seq from mouse cortex from ENCODE, Rad21 ChIP-Seq from *Rad21*$^{Lox/Lox}$ (*L/L*_Rad21), CTCF-, Stag1, Stag2-depleted (*Δ/Δ*+siRNA_Rad21, aligned with the respective sample) and control *Rad21*$^{Δ/Δ}$ ASTs (*Δ/Δ*_Rad21). White squares highlight the 1.5 Mb chromatin domain described.

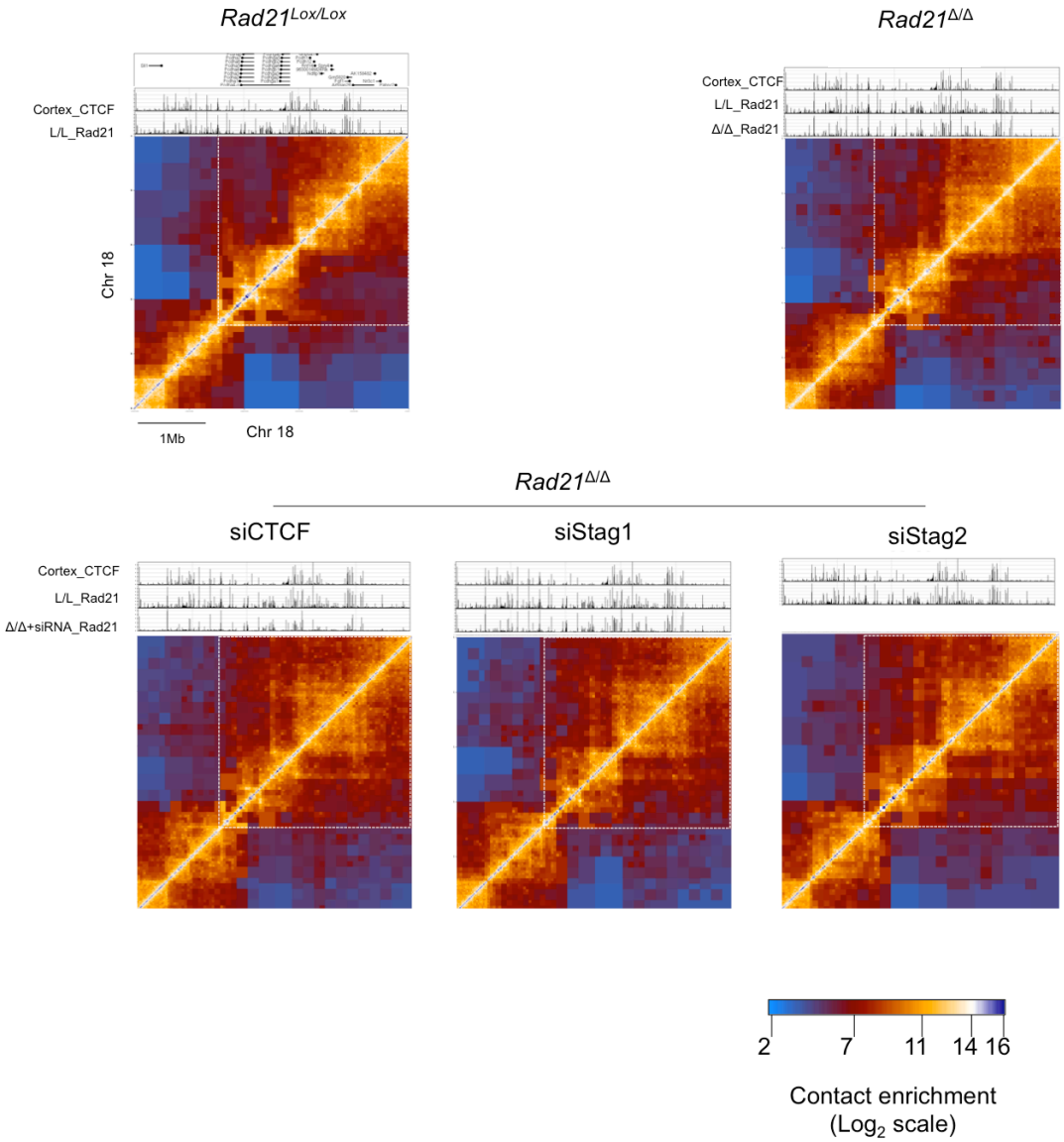

Figure 39. CTCF and Stag depletion lead to ectopic chromatin contacts from the cohesin/CTCF-regulated Pcdh cluster to the neighbouring domains in cohesin-deficient ASTs. Hi-C contact maps of a wide 5 kb region on chromosome 18, encompassing the Pcdh cluster of genes, in *Rad21*$^{Lox/Lox}$ and in CTCF-, Stag1, Stag2-depleted, control *Rad21*$^{Δ/Δ}$ ASTs. Aligned are the genes spanning the region (top panel in *Rad21*$^{Lox/Lox}$), CTCF and Rad21 ChIP-Seq signal (top 1%) in the samples indicated on the left side of the tracks. CTCF ChIP-Seq from mouse cortex from ENCODE, Rad21 ChIP-Seq from *Rad21*$^{Lox/Lox}$ (*L/L*_Rad21), CTCF-, Stag1, Stag2-depleted (*Δ/Δ*+siRNA_Rad21, aligned with the respective sample) and control *Rad21*$^{Δ/Δ}$ ASTs (*Δ/Δ*_Rad21). White squares highlight the two neighbouring chromatin domain of interest.

CTCF DEPLETION CORRELATES WITH INCREASED CELL DEATH IN COHESIN-DEFICIENT ASTS

Depletion of Rad21 in cycling NSCs leads to cell death, however post-mitotic ASTs do not experience cell death upon Rad21 depletion (described in CHAPTER 2). I explored whether this was due to the residual cohesin protein in post-mitotic ASTs, and examined the status of cell survival once ASTs were treated with siCTCF when levels of Rad21 had dropped further.

I observed that upon transfection with siRNAs, CTCF-depleted *Rad21*$^{\Delta/\Delta}$ ASTs had significantly more cells floating in the media compared to their respective controls. So to quantify potential changes in cell viability or cell cycle, PI staining and cell cycle analysis in cell pellets fixed with EtOH was performed. Using this method, not only cells found in the different cell cycle phases are quantified but also dead and dying cells are found in the sub-G_1 fraction. This method is not able to distinguish between different cell death pathways and therefore represents a broader quantification of cell death.

I performed the analyses using the adherent cultures including the supernatants, which represented the whole population including dying or dead cells. The results were obtained from two timepoints (48 hours and 75 hour) post transfections with siRNA-mediated knock-downs (siNTR and siCTCF). Since ASTs are not cycling, NSCs were also included as a control for the investigation of potential cell cycle effects. Initially, I didn't observe any changes in sub-G_1 population transfected with siNTR compared to untransfected control ASTs. This result suggests that the transfection doesn't lead to an increase in cell death. However, I observed a 2-fold increase in the sub-G_1 population in Rad21$^{\Delta/\Delta}$ ASTs 48 hours after siRNA transfection for CTCF, implying that CTCF depletion leads to a decrease in cell viability already at 48 hours post transfections (5% to 10% Sub-G_1 population). A similar increase in cell death (10% to 20% Sub-G_1 population) was again noted at 75 hours post transfection with siCTCF in *Rad21*$^{\Delta/\Delta}$ ASTs (**Figure 40**). This analysis also showed that there was no cell cycle re-entry due to transfections or RNAi-mediated depletion.

Thus, CTCF depletion in cohesin-deficient ASTs causes an increase in cell death. However, it should be noted that enrichment of the viable and non-cycling cells were collected for all the experiments (i.e. ChIP-Seq, Hi-C and protein analysis) without supernatants so that a degradation of the genomic or protein material would not interfere with the interpretation of the results.

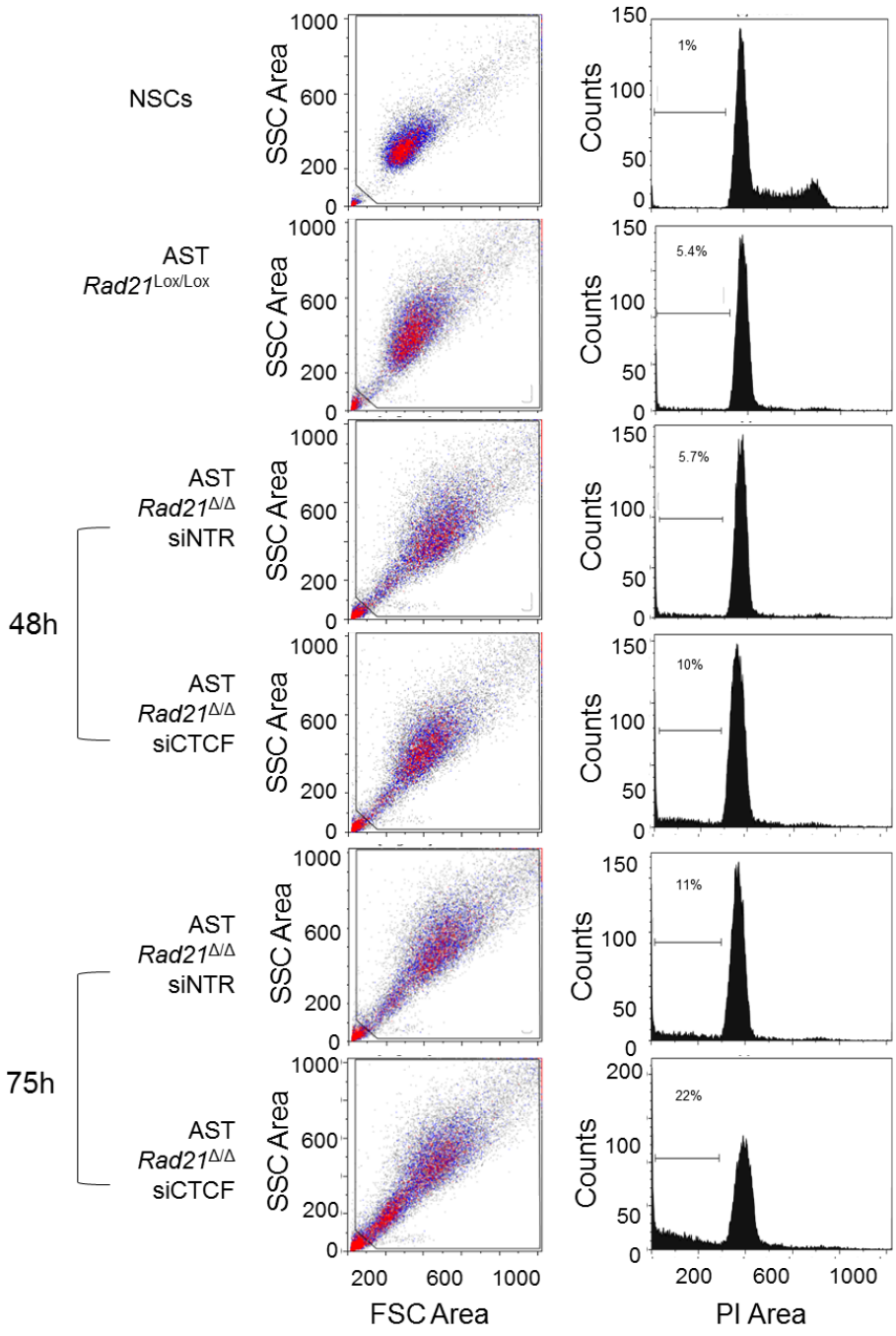

Figure 40. CTCF depletion causes increased cell death in Rad21-deficient ASTs. Flow cytometry results in left panels depicting the side scatter (i.e. conformation of the inner morphology) to the forward scatter (i.e. cell diameter) areas and cell cycle analysis after PI staining in Rad21$^{Lox/Lox}$ ASTs and NSCs used as a control for cell cycle changes and *Rad21*$^{\Delta/\Delta}$ ASTs transfected with siRNA for CTCF and non-targeting control (NTR) for 48 and 75 hours (10,000 events per sample). (Sub-G1%; indicated above).

3.3. DISCUSSION

In CHAPTER 2, it was shown that genetic deletion of Rad21 led to chromosomal domain decompaction, loss of specific long-range contacts and subsequent gene deregulation. However, the gene expression and genome structure changes were subtle and the cells were viable. I explored the hypothesis that the existing cohesin complexes in Rad21-deficient ASTs are sufficient to maintain chromatin structure. I tested how the destabilisation of residual cohesin would influence chromatin organsation.

Residual chromatin-bound cohesin complexes were present in *Rad21*$^{\Delta/\Delta}$ post-mitotic ASTs. Residual cohesin was enriched for known regulators of cohesin association with chromatin. Upon DNA replication, some of the regulators stabilise cohesin on chromatin allowing SCC (i.e. Smc3 acetylation) (Beckouet et al., 2010; Ben-Shahar et al., 2008). This type of 'cohesed' cohesin is proposed to be the 'stable' pool with a time of residence of 6 hours in G_2 mammalian cells (Gerlich et al., 2006) and to topologically embrace sister chromatids in yeast (Gligoris et al., 2014). I hypothesise that by depleting Rad21, I enriched for a more 'stable' cohesin pool that has lower turn-over on chromatin. This hypothesis is supported by the fact that Rad21 mRNA levels were almost undetectable. Furthermore, the 4C-seq timecourse experiments (discussed in CHAPTER 2) also support this hypothesis. Other than stability, the residual cohesin complexes could be enriched for complexes, which associate with chromatin as this proposed by the handcuff model (discussed in CHAPTER 1). The model would predict that two cohesin rings entrap two different DNA strands and as they are associated via Stag and Pds5 proteins, they create a chromatin loop. However, this doesn't exclude the possibility that stability and different modes of cohesin association co-exist in Rad21-deficient cells.

I hypothesised that the depletion of cohesin regulators would destabilise cohesin in Rad21-depleted cells. Indeed, preliminary results showed that the depletion of selected regulators led to further decrease in cohesin chromatin levels as well as at selected cohesin binding sites across the genome and consequently, to loss of cohesin-anchored contacts globally. Interestingly, we noted a different sensitivity of cohesin-mediated insulation upon depletion of the different factors, with CTCF and Stag1 having the most prominent effects. These results were supported by the decreased number of cohesin binding sites in CTCF RNAi-depleted (Schmidt et al., 2010) or Stag1-null (Remeseiro et al., 2012b) mammalian cells. However, one should bear in

mind that the above studies target the total cohesin population whereas we target the residual cohesin pool after Rad21 genetic depletion.

Given that CTCF and cohesin are directly interacting through Stag1 or Stag2 (Xiao et al., 2011), one would expect an additive effect after CTCF depletion in cohesin chromatin binding and chromatin structure. The difference in the effect from Stag2 compared to Stag1 on the Hi-C analysis that I reported could be explained by that they have different basal levels in ASTs and after the similar decrease they experience in Rad21-deficient cells, one of them could be in very low levels on chromatin. So its further depletion wouldn't add an effect in our experiments (discussed in CHAPTER 4). Another possibility is that the interface between CTCF and Stag proteins differs between Stag1 and Stag2 (e.g. protein modifications) and in turn, confer different stabilities and/or functions to the cohesin complex. Showing that depleting Stag1 didn't affect Stag2 levels and vice versa, I propose that residual Stag1-cohesin complexes are more essential to cohesin-mediated domain structure than Stag2 is in this post-mitotic system.

Cells treated with siCTCF experience changes at Rad21 binding sites as well as a loss of contact insulation from changing sites. These changes included loss of internal domain structure and new long-range interactions whereby domains became 'merged'. siCTCF cells also experienced an increase in cell death compared to control cells which may arise by the low levels of cohesin reached only by siCTCF. This implies that there might be a threshold of cohesin levels necessary for the maintenance of chromatin structure and cell viability in ASTs. There is evidence in the literature of such examples. In CdLS, mutation in the cohesin loading factors cause 30% decrease in its protein levels but cells do manage to survive and create a human being (Kawauchi et al., 2009). While the cell death phenotype is unlikely due to ASTs re-entering the cell cycle, it may be a function of the structural changes we observed, and it may also be linked to specific changes in transcription which the cells are very sensitive to.

In this context, I excluded the effects of CTCF-independent cohesin functions in chromatin architecture. Several evidence point out that cohesin facilitates CTCF insulator functions. Sofueva *et al.* showed that CTCF sites not co-localising with cohesin show minimal contact insulation profiles at multiple scales. Therefore, if there is an effect from CTCF sites, this wouldn't be able to completely account for the changes we see here. Additional evidence suggest that the minimal insulation seen in CTCF-only sites is due to the presence of cohesin/CTCF sites near them and therefore

influencing the chromatin folding around the area (M. Vietri Rudan, unpublished data). I proposed that the effect I observed by Hi-C is due to the role of CTCF in maintaining cohesin stability on chromatin and therefore its functions in chromatin architecture.

Taken together with previous data, I have showed that the decrease in cohesin levels on chromatin were accompanied with decreased cohesin binding and loss of cohesin-anchored contacts that shape chromatin domains. Here, cohesin destabilisation off chromatin was achieved by depleting a series of cohesin regulators and the results highlighted different sensitivity to the different factors. This finding implied that cohesin may be stabilised by different mechanisms on chromatin and that these mechanisms in turn mediate different functional chromatin contacts. Additional high-resolution 4C-Seq experiments from different classes of cohesin sites would aid us to dissect further these different mechanisms.

CHAPTER 4.

DISCUSSION

CHAPTER 4. DISCUSSION

A closer look of genome organisation and gene regulation showed that chromatin is not organised randomly and it is interlinked to its functions (Cremer and Cremer, 2001; de Wit and de Laat, 2012). Factors which were central to such studies are cohesin and CTCF proteins, which have essential roles in mediating gene regulation *in cis* (Hadjur et al., 2009; Kagey et al., 2010; Nativio et al., 2009; Seitan et al., 2011). The field progressed from descriptive studies (Dixon et al., 2012; Sexton et al., 2012) to functional studies asking about the role of chromatin domains (Lupianez et al., 2015; Nora et al., 2012) and proteins, such as cohesin, in these.

First, we showed that the cohesin complex has a global role in chromatin domain maintenance forming the basis for gene expression. Cohesin-deficient post-mitotic ASTs experienced a genome-wide perturbation of transcription and loss of specific cohesin-anchored chromatin contacts leading to chromatin domain decompaction. These changes were accompanied by a significant increase in nuclear volume and perturbed nuclear morphology.

Subsequently, we showed a role for cohesin regulators in cohesin stabilisation on chromatin and therefore, in cohesin-mediated chromatin domain maintenance. The depletion of CTCF and Stag proteins resulted in a further decrease in Rad21 levels and binding on chromatin which were correlated with a change in chromatin domain border definition. Importantly, this led to a decrease in cell viability in cohesin-deficient post-mitotic cells.

4.1. COHESIN'S ROLES IN THE POSTMITOTIC NUCLEUS

COHESIN DOSAGE AND COHESIN FUNCTIONS

Cohesin proteins have critical roles in DNA replication and chromosome segregation, both during mitosis and meiosis (Hirano, 2006; Merkenschlager and Odom, 2013; Nasmyth and Haering, 2009; Peters et al., 2008). Initial studies and our findings suggest that different cohesin dosage may play a role in the different cohesin functions.

Cohesin is expressed in post-mitotic cells and tissues, such as mouse cortex (Remeseiro et al., 2012b; Wendt et al., 2008), which is supported by my observations that cohesin subunits are expressed in post-mitotic ASTs. Interestingly, Rad21 mRNA

and protein expression was decreased in ASTs to roughly half of the amount in NSCs. This might be explained by that fact that ASTs don't require cohesin's chromosome segregation functions and only require cohesin's gene regulation functions. All these findings imply that different cohesin levels may be required for its different functions. In yeast, a study where cohesin amounts were titrated showed that different amounts are required for a group of cohesin functions they investigated (Heidinger-Pauli et al., 2010). 30% of the original amount of cohesin in yeast is enough to maintain its function in cohesion but not those in DNA repair, making cells sensitive to double-strand break reagents. For cohesion to be impaired, cohesin levels needed to be further decreased to 13%. In humans, an example is CdL syndromes, where lower cohesin dosage doesn't cause cohesion defects but alters gene expression profiles in CdLS mouse models and patient samples (Fullwood et al., 2009a; Liu and Krantz, 2009). In these examples, the decrease in cohesin levels was not sufficient to compromise the growth of an embryo but instead led to defects in its development. Therefore, it might be possible that only mild changes in cohesin levels are required for its gene regulator function whereas more dramatic or complete ablation of its levels leads to SCC defects and cell death. These thresholds could be different in different species (e.g. yeast versus mammals) or even different tissues.

Which are the mechanism(s) involved in regulating cohesin dosage? Together with the decrease in mRNA levels of the cohesin genes upon differentiation of NSCs to ASTs, I observed a decrease in expression of cell cycle regulated genes by RNA-Seq and qPCR. This set of genes (e.g. cyclin D1) is responsible for the initiating events which promote the G_1-to-S transition. One mechanism which regulates such events is the phosphorylation of retinoblastoma protein (pRb) which, prior to its phosphorylation, inhibits E2F-mediated transcription (Dimova and Dyson, 2005). E2F transcription factor targets the transcription of genes involved in DNA replication and in turn, is fundamental for the transition of the cell from G_1-to-S transition. In the transcriptional analysis by RNA-Seq, many of the E2F target-genes are down-regulated upon differentiation of NSCs to ASTs (e.g. Mini-Chromosome Maitenance, MCM, proteins). Could it be that cohesin gene dosage is regulated by such cell cycle 'sensing' mechanisms? If I was to investigate this further, I would analyse the binding profiles of factors, such as E2F, around the transcriptional unit of cohesin genes, such as Rad21, by ChIP methods and quantify cohesin mRNA levels as cells exit mitosis. Evidence about the regulation of cohesin gene expression are lacking so this approach would provide insight into the presence or absence of such cell cycle 'sensing' factors in the

regulation of cohesin genes. If this is true, it would be essential to our knowledge on cohesin roles in development and disease.

COHESIN AND THE MORPHOLOGY OF THE POST-MITOTIC NUCLEUS

Studies looking at the nuclear volumes in cells from CdLS patients or cells expressing cohesin mutations (Gard et al., 2009; Nolen et al., 2013) suggest a role for cohesin proteins in nuclear size and morphology. When I decreased cohesin levels in cohesin-deficient post-mitotic cells, I reported a similar effect on nuclear volume and changes in nuclear morphology. Nuclear size and morphology can arise due to changes in the inner nuclear structures, such as the nuclear lamina layer (Goldman et al., 2004). Nuclear lamina is linked to chromatin organisation and transcriptional activity as well as the tension mechanisms that hold the nuclei together (Webster et al., 2009). I hypothesise that changes in cohesin levels may influence inner nuclear structures and interfere with their association with the lamina. Indeed, a study on nuclear lamina components and their interaction with the genome in human fibroblasts, showed that CTCF is enriched at the borders of lamina-associated domains (LADs) (Guelen et al., 2008). So it was proposed that CTCF might have a role in maintaining the LADs. If that is true, it is possible that cohesin is also involved since it is known to facilitate CTCF insulator function. Therefore, cohesin depletion may disrupt the lamina-chromatin association. One could expect that if these cells were challenged further (i.e. mechanical stress or cell cycle re-entry), they would not be able to maintain their viability since their nucleus is already compromised, possibly leading to the phenotypes described herein.

COHESIN AND CTCF ARE KEY ELEMENTS OF GENOME ORGANISATION

It has been suggested that TADs are functionally important, even during development, evolution and disease (Lupianez et al., 2015; Nora et al., 2012; Vietri Rudan et al., 2015). Several important questions remain. Which proteins are involved in forming and/or maintaining TADs? Through the study of cohesin-deficient post-mitotic cells, cohesin with CTCF were shown to have a functional role in the regulation of chromatin contacts that partition the genome. Cohesin depletion in post-mitotic cells leads to global topological domain decompaction with decreased contacts within and increased contacts between TADs (Sofueva et al., 2013). The decreased contacts within TADs

and global gene de-regulation agree with studies in non-cycling or cycling mammalian cells, with either cohesin genetic depletion (Seitan et al., 2013) or TEV-mediated cleavage (Zuin et al., 2014), respectively. Support to our phenotype in cohesin-deficient ASTs also comes from a recent study, which combined 5C and mathematical modeling with single-cell methods to predict different conformations of the chromatin fibers at the X-inactivation centre. The authors 'deleted' *in silico* a CTCF/cohesin binding site located internally to a TAD from their model and saw both a decrease in intra-domain contacts and an increase in inter-domain contacts (Giorgetti et al., 2014). This candidate *in silico* approach also supports a cohesin/CTCF role in TAD structure maintenance.

However, there are differences in the models proposed by the above studies. Zuin *et al.* proposed a role for CTCF in maintaining TAD structure, and for cohesin in maintaining their internal chromatin contacts (Zuin et al., 2014). However, Zuin *et al.* study doesn't accommodate for cell death caused by cohesin depletion in cycling cells or differences between the siRNA-mediated CTCF depletion and TEV-mediated cleavage of cohesin. The proposed model, where CTCF has different roles than cohesin, is not supported by our data since CTCF-only sites were shown to engage in minimal contact insulation (Sofueva et al., 2013). In our study, contact insulation was undetectable when it was calculated from CTCF-only sites located 50kb away from other CTCF/cohesin sites (M.Vietri Rudan; unpublished data). So I propose that CTCF and cohesin are both required to mediate genome organisation.

However, the demarcation of the TADs was still identifiable in the studies discussed above, which suggests that cohesin may not be involved in TAD maintenance or that additional factors are required. In *Drosophila*, there is a group of proteins involved in insulation and implicated in chromatin domain structure. It was proposed that the number of insulators is correlated to the strength of the border of the TADs (Van Bortle et al., 2014). This simplistic view explaining border strength probably does not reflect the complexity of genome organisation but it does give an indication that multiple proteins may be involved in TAD maintenance. This hypothesis would predict that depleting one protein at a time wouldn't be suffient to disrupt chromatin domains. In agreement with this, I reported that the most pronounced effect on global contact insulation profiles resulted from CTCF/Rad21 depletion than with Rad21-alone. Therefore, it is conceivable that more than one protein may be required to mediate chromatin domains at the TAD level. But which proteins could be involved, if not cohesin proteins, in holding the TAD structures. Alternative candidates could be other

members of the SMC family of proteins, such as condensins. In *C. elegans*, where CTCF is not present, the dosage-compensation complex (DCC) was proposed to have direct roles in the maintenance of the TADs on the X-chromosome (Crane et al., 2015). DCC is able to do this by localising to specific sites, called Rex, which are enriched at the border regions of the TADs. Therefore, TADs may require additional factors for their maintenance.

Another possibility could be that cohesin and CTCF are required to *establish* formation of TADs as cells exit mitosis, and are not required for their *maintenance*. Our study and others wouldn't be able to distinguise the above. What if cohesin and CTCF demarcate the sites where the domains first emerge in G_1? A similar idea is gene bookmarking according to which binding of certain transcription factors is maintained at promoters as cells exit mitosis. This mechanism is essential for cell identity (Zaidi et al., 2010). So cohesin, and more likely CTCF, could be 'bookmarking' the formation of TADs. Given that CTCF recognises a specific sequence motif, one which is encoded directly in the DNA sequence, CTCF could be a plausible candidate for 'TAD bookmarking'. A recent study comes to argue for a CTCF role in mitotic bookmarking in *Drosophila.* The authors found a class of CTCF sites, which are maintained through interphase and mitosis (i.e. MI sites), frequently enriched at TADs borders (Shen et al., 2015). If we wanted to explore a more functional role of CTCF in 'bookmarking' the formation of TADs, we would first perform ChIP-Seq experiments to investigate the sites in the genome where CTCF is enriched as cells exit mitosis and then, delete these CTCF sites to evaluate their functional significance in chromatin structure by using a combination of CRISPR and 3C methods. Using this approach, I would be able to investigate the roles of either cohesin or CTCF in 'bookmarking' the formation of chromatin domains and provide direct evidence as to their roles in genome organisation as it emerges in G_1.

Taken together, these functional studies suggest an important role for cohesin and CTCF in TAD maintenance providing insights to its functions. However, it remains unclear whether cohesin and CTCF have a causal relationship with the maintenance and or establishment of the TADs.

RESIDUAL COHESIN COMPLEXES

Residual cohesin complexes were present on chromatin in cohesin-deficient ASTs. What is the reason why some cohesin complexes remain on chromatin upon Rad21 genetic deletion? Is it possible that I enriched for stable complexes in cohesin-deficient ASTs with longer residence time on chromatin? Our data indicate that residual cohesin complexes are enriched for Smc3 acetylation. This modification is known to mark cohesin complexes during SCC establishment in S-phase (Beckouet et al., 2010). One way to investigate cohesin stability on chromatin is by using FRAP after introducing a GFP-tag in the endogenous floxed *Rad21* locus. This way we can investigate whether the residual cohesin has increased time of residence in chromatin compared to the cohesin population in control ASTs. An alternative way to test whether specific sites in the genome are more stable than others uses a variation of the ChIP method, namely Cross-Linking Kinetic (CLK) (Poorey et al., 2013). In yeast, a timecourse of formaldehyde fixation between 1 second and 30 minutes and ChIP-qPCR experiments were used to predict the kinetics of DNA-binding proteins on chromatin at specific sites. The principle is that when formaldehyde is added to the cell cross-linking occurs rapidly and therefore protein occupancy is captured so the off rate of crosslinked proteins is eliminated. If the time that formaldehyde is present increases, more protein binding events will be captured. However, the timepoints need to be in a lesser time of the off rate of the protein so that the kinetics are captured. Using these methods, I would be able to explore whether the residual cohesin complexes are more stably bound to chromatin in ASTs.

Is it possible that cohesin is associated with chromatin in a different way at the sites where it remains in Rad21-depleted ASTs? In yeast cells, cohesin has been shown to entrap single DNA fibers and replicated DNA (i.e. two DNA fibers), namely with the 'embrace' model (Huis in 't Veld et al., 2014). In mammalian cells, another model, namely the 'handcuff' model, which predicts that a single chromatid is entrapped by a single cohesin ring, and that the two cohesin rings then interact with one another to connect the two sister chromatids (Huang et al. 2005; Zhang et al. 2008b). This could occur via one molecule of Pds5A/B and one of Stag1/2 proteins connecting the two kleisin subunits (Zhang et al., 2013) and therefore, hold the cohesin 'handcuff' together. However, the evidence supporting the presence of either model doesn't exclude their combination on mammalian chromosomes. If both modes exist in ASTs, they may or not have different stabilities on chromatin. In this respect, one might hypothesise that the depletion of Pds5 or Stag proteins would compromise the

connection between the two rings in a 'handcuff' conformation but without removal of the cohesin ring off chromatin. In support of this argument, a study in yeast shows that either Stag or Pds5 depletion lead to SCC problems but not to cohesin removal off chromosomes (Kulemzina et al., 2012). This becomes central if you consider that different combinations of the components, such as Stag1-Pds5A, might themselves confer different stabilities to the complex (discussed below).

STAG2, STAG1 AND CTCF DEPLETION LEAD TO A PROGRESSIVE LOSS IN CHROMATIN STRUCTURE

The depletion of selected cohesin regulators led to further destabilisation of residual cohesin off chromatin and in turn, changes in chromosomal domain structures. I showed that CTCF and Stag1 depletion have a bigger effect on chromatin domains and contact insulation profiles around cohesin/CTCF sites than Stag2 in Rad21-deficient ASTs.

Why did I observe this difference between Stag1 and Stag2 in cohesin-deficient ASTs? Given that their levels are equally affected in Rad21-deficient ASTs and that Stag1 doesn't compensate for Stag2 upon its depletion and vice versa, one possible explanation for the difference I observed in the contact insulation profiles (i.e. Stag1 depletion influenced contact insulation more than Stag2) is that the absolute levels of Stag1 and Stag2 protein are different in ASTs. Indeed, it has been shown that the levels of Stag1 and Stag2 proteins are different in some tissues, and that the proportion varies depending on the tissue type (Canudas and Smith, 2009; Remeseiro et al., 2012b). If we wanted to test this hypothesis, we would use quantitate mass spectrometry analysis after immunoprecipitation of the cohesin complexes for robust quantification of the Stag1/2 relative amount in ASTs.

Another explanation for the differences between Stag1 and Stag2 could be that the interface of CTCF-Stag is different between Stag1- and Stag2-cohesin complexes. While the homology of the sequence between Stag1 and Stag2 is 70%, the 75 amino acids of their N-terminal region is divergent between them (Xiao et al., 2011). It is with their N-terminal region that Stag proteins associate with the C-terminal region of CTCF. This difference in sequence of Stag1 and Stag2, where they associate with CTCF, may influence the stability of their interaction with CTCF and therefore, that of the cohesin complex at CTCF sites on chromatin.

Post-translational modifications on Stag1/2 and/or CTCF might change their conformation or their susceptibility to other cohesin regulators. This could have an impact on either cohesin stability or function. The CTCF C-terminal region, is where the CTCF-Stag association is reported to occur (Xiao et al., 2011) and this region is known to undergo phosphorylation on CTCF. This modification has been reported to modulate CTCF function as an insulator without interfering with its binding to DNA (El-Kady and Klenova, 2005; Klenova et al., 2001). This modification could lead to de-stabilisation or ablation of Stag1/2-CTCF interaction and in turn, have a functional impact on CTCF/cohesin roles in mediating chromatin loops; whether this modification has a different impact on Stag1 versus Stag2 is yet to be clarified.

Also there are post-translational modifications on Stag proteins, which are shown to be important for SCC and cohesin removal during the 'prophase pathway'. PLK-mediated phosphorylation of the C-terminal domain of Stag1/2 decrease cohesin association with DNA (Hauf et al., 2005; Sumara et al., 2002). These modifications have been are characterised for cohesin's roles in mitosis, but as we found several components of cohesin regulation during cell cycle expressed in post-mitotic ASTs (e.g. Esco1), it wouldn't be surprising to find also the modifications on Stag proteins.

While the modifications (e.g. phosphorylation) on Stag proteins are not on the region where it interacts with CTCF, they might be changing its conformation and therefore impacting its stability or interactions with other factors, such as Wapl. Thus, this interface could be the target of several proteins and modification that confers different stabilities and possibly different functions to the Stag1- and Stag2-cohesin/CTCF complexes. If this is true, given that the Pds5A and B can associate with either Stag1 and 2 (Losada et al., 2005), they increase the diversity of cohesin complexes and their stability on chromatin. In this context, if cohesin rings were associated with chromatin with the handcuff model, they could be more sensitive to changing levels of the factors described above or to their modifications, either Pds5 or Stag proteins. Indeed, Pds5 or Stag depletion in yeast leads to loss of cohesion but not to a decrease in cohesin levels on chromatin (Kulemzina et al., 2012), implying that the depletion of selected cohesin regulators may disrupt its functions but not its association/presence with/on chromatin. Whether this finding is due to the depletion of the proteins holding the 'handcuff' together (i.e. Pds5, Stag) and how this might be influenced by the depletion of the different paralogues (i.e. Stag1 versus Stag2), is yet to be clarified. Since cohesin functions are linked to how it associates with chromatin, it is vital to understand deeply the modes of cohesin association and the stability these may confer.

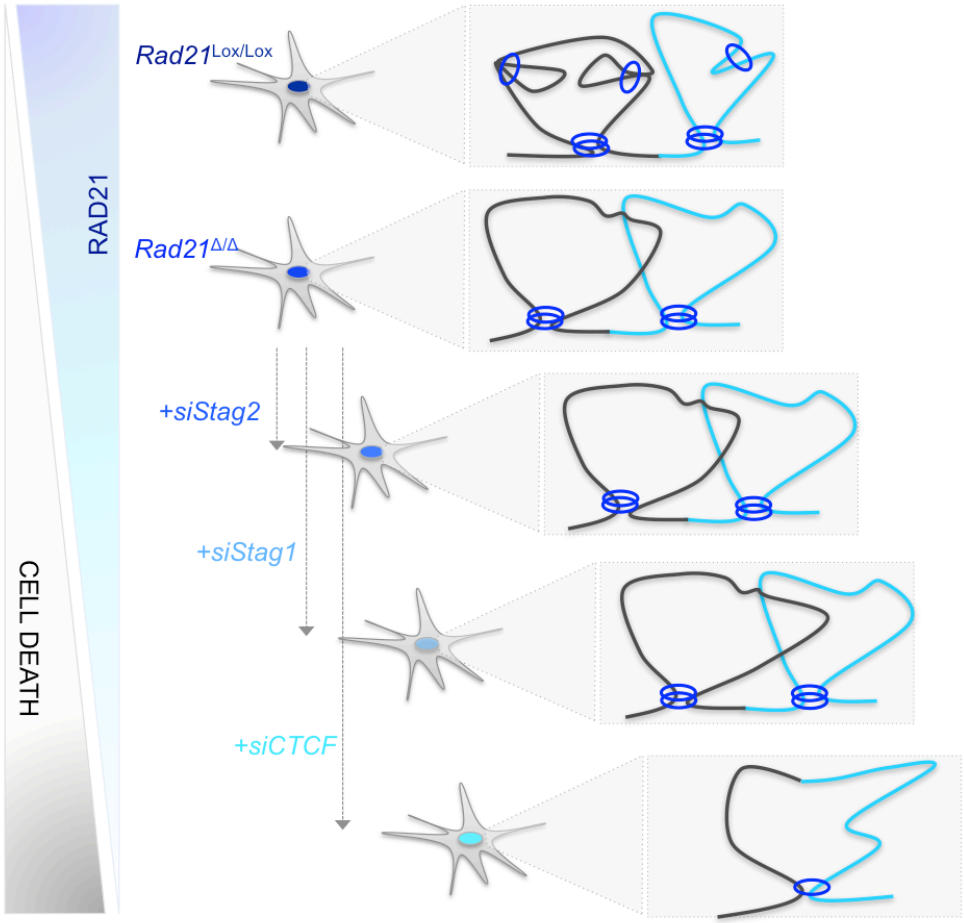

Figure 41. Stag2, Stag1 and CTCF depletion lead to a progressive loss in chromatin structure following the decreasing cohesin levels on chromatin. Graphical representation of the proposed model where cohesin (i.e. Rad21) decreasing levels (blue triangle) due to the knock-down of the cohesin regulators (i.e. Stag1/2 and CTCF) in *Rad21*$^{\Delta/\Delta}$ ASTs leads to a further decompaction of topological domains and an increase in cell death (grey triangle). In ASTs, two neighbouring domains are depicted (grey and blue) and cohesin (dark blue) is mediating both large scale domain loops as well as smaller, internal domain loops. After Rad21 genetic deletion, domains were decompacted but domain borders remained intact in *Rad21*$^{\Delta/\Delta}$ ASTs. The additional depletion via RNAi methods of the main cohesin regulators led to further decrease in cohesin levels on chromatin, which resulted in the loss of domain borders and cell death in the case of CTCF knock-down. This evidence come to support a key role for cohesin proteins in genome organisation in a post-mitotic model system where the roles of cohesin in sister chromatid cohesion were eliminated.

CTCF depletion had the strongest effect in cohesin protein levels, binding and contact insulation profiles. This effect in protein levels and contact insulation resembles the sum of the two Stag1 and Stag2 together. These findings underline the role of CTCF and cohesin in global genome organisation. Although CTCF has been suggested to mediate contact insulation independent of cohesin, as indicated by studies in cycling cells (Phillips-Cremins et al., 2013; Wendt et al., 2008) Phillips-Cremins et al. 2013), our data don't agree with this model (discussed above). Further, it would be interesting to investigate how chromosome structure is affected in Esco1- in comparison to CTCF-depleted Rad21-deficient ASTs. Like CTCF, in Esco1-depleted cells both Stag1 and Stag2 complexes were reduced. The fact that cohesin levels were also decreased upon Esco1 depletion could mean that Smc3 acetylation is involved in maintaining cohesin on chromatin even in post-mitotic cells.

The depletion of CTCF led to a further decrease in Rad21 protein levels on chromatin. This results are supported by the decreased cohesin binding sites in CTCF RNAi-depleted (Schmidt et al., 2010) or Stag1-null (Remeseiro et al., 2012b) mammalian cells. It would be interesting to quantify the changes in cohesin binding profiles with those in chromatin structure to show their correlation. For this, we need to quantify robustly the changes in ChIP signal between the different conditions, so additional strategies would be useful. One could be to include either an exogenous DNA control throughout the protocol (Orlando et al., 2014) or treat all libraries simultaneously by multiplexing them before the ChIP enrichment step (Lara-Astiaso et al., 2014).

Surprisingly, CTCF depletion in cohesin-deficient ASTs decreased cell viability. The changes in chromatin organisation cannot be directly linked to the cell death phenotype because the methods used were not sufficient to exclude the CTCF-dependent deregulation of single genes important for cell viability. In order to investigate this further a dissection of the transcriptional profiles of these cells through RNA-seq would be necessary. Taken together, the decrease in cohesin levels, the decrease in contact insulation and the increased cell death we observed in CTCF-depleted Rad21-deficient ASTs support the argument that different cohesin amounts are essential for different cohesin functions. The low amounts of cohesin reached with CTCF depletion could be the threshold for cell viability in this post-mitotic system.

In an elegant approach combining 5C, modeling and single-cell methods, it was proposed that chromatin loops within TADs may be responsible for stability of the TAD

borders (Giorgetti et al., 2014). In this context, it is conceivable that cohesin and CTCF together stabilise chromatin contacts that otherwise engage in random movements in the nucleus. At TAD borders, more stable interactions between CTCF/cohesin sites might participate in the definition of large-scale chromatin domains within which functional interactions happen; within a domain, CTCF/cohesin might help define more dynamic gene-regulatory loops. This spectrum of stability might be essential for the role of cohesin in genome functionality and it might rely on a diverse range of regulation (e.g. post-translational modifications), composition of the complexes (e.g. Stag1/2) and modes of association of the cohesin rings with chromatin (e.g. 'embrace' vs 'hancuff' models).

4.2. CONCLUSION AND FUTURE PERSPECTIVES

It is clear that the structure of the genome is interlinked to its functions in propagating and expressing genetic information. All these processes are crucial for a single cell and its fate to the development of an organism and the evolution of the species. Disruption of nuclear lamina by a mutation of a single gene or genome architecture even at localised sites had dramatic effects in human development and disease. But is yet to be elucidated which are the mechanisms by which key protein players are mediating the interplay between chromatin structure and functions with implications in development and disease. In this thesis, I showed that cohesin is one of those key factors with functional roles in global chromatin domain maintenance and gene regulation while suggesting different mechanisms by which cohesin and chromatin interaction are stabilsed in post-mitotic nuclei with implications in cell viability. Hence, these and future findings in this context would be fundamental to our understanding of how the genome folds and functions in a way that allows or prevents cell survival, development and disease.

CHAPTER 5.

METHODOLOGY

CHAPTER 5. METHODOLOGY

5.1. MAMMALIAN CELL CULTURE

IN VITRO ADHERENT CULTURES OF MOUSE NSCS

The mouse embryonic stem cells (ESCs) contained LoxP sites flanking the fifth and sixth exon of *Rad21*, the core cohesin subunit gene, on both alleles and an insertion of the estrogen receptor (ER) fused to Cre recombinase (ERT2-Cre system) within the Rosa locus (herein reffered as ESCs $Rad21^{Lox/Lox}$). When Cre-Recombinase was expressed, after ER substrate delivery (i.e. 4-hydroxytamoxifen addition, OHT), it mediated recombination between the LoxP sites, which resulted in the deletion of the flanking sequences. Adherent neural stem cells (NSCs) were derived from ESCs and clones were collected and expanded previously by Dr S.Hadjur according to the protocol in (Conti and Cattaneo, 2010; Conti et al., 2005) (herein reffered as NSCs $Rad21^{Lox/Lox}$). As a control, NSCs were obtained, which were wild type for Rad21 gene while they were expressing the fusion protein (NSCs $Rad21^{WT/WT}$). Mouse NSC cultures were maintained as adherent monolayers in condition media DMEM/Ham's F-12 media with supplements [For 500 ml DMEM/Ham's F-12 (Sigma): 7.25 ml D-glucose (2 M, SIGMA), 5 ml Non-Essential Amino Acids (100x, PAA), 0.5 ml 2-mercaptoethanol (50 mM, Invitrogen), 0.8 ml BSA solution (75 mM, Invitrogen), 5 ml B27 supplement (Invitrogen), N2 (Invitrogen)] and murine epidermal growth factor (mEGF, Peprotech), human fibroblast growth factor 2 (hFGF-2, Peprotech), and Laminin from Engelbreth-Holm-Swarm murine sarcoma basement membrane (Sigma) at final concentration of 10 ng/ml. As described in (Conti et al., 2005), the extrinsic factors EGF and FGF-2 are sufficient to sustain symmetrical self-renewing divisions of NSCs. NSC cultures were expanded by using accutase (Sigma) for 5 minutes at room temperature followed by a wash with 10 ml of NSC media and centrifugation at 1,200 rpm for 4 minutes.

IN VITRO DIFFERENTIATION OF MOUSE NSCS INTO POST-MITOTIC ASTS

Post-mitotic ASTs were obtained by the differentiation of NSCs so that the roles of cohesin proteins in G_1/G_0 were investigated while the cell cycle bias in eliminated. NSCs were plated in Laminin-coated plates (10 ng/ml in PBS; 2 hours to overnight at 37^{o}C) so that they reached 60% confluence by 24 hours. Then, they were treated with

10 ng/ml of human bone morphogenetic protein 4 (BMP4, R&D Systems) while EGF was withdrawn from the media and the concentration of FGF was decreased to 1 ng/ml. ASTs were maintained as adherent cultures until they were harvested for downstream experiments. In the case that the re-plating of ASTs was required for down-stream experiments, ASTs were plated by using accutase (PAA) for 10-15 minutes at room temperature followed by a wash with 10 ml of AST media and centrifugation at 1,200 rpm for 4 minutes.

RAD21 COHESIN SUBUNIT DELETION FROM NSCS AND ASTS

Rad21 depletion was necessary and sufficient in both cell types in order to investigate its impact both in dividing and post-mitotic cells in different experimental settings (Hadjur et al., 2009). In NSCs, *Rad21* was depleted by the addition of 200 nM OHT to the condition NSC media for 48 hours. In ASTs, NSCs were initially platted with respect to optimal differentiation conditions (i.e. 60% confluency) and after 24 hours, they were differentiated into ASTs in the presence of BMP4. After an additional 24 hours, *Rad21* was depleted by infections with 10^7 plaque-forming unit (PFU) of Adenovirus-CMV-Cre (Cre-Ad, Vector Biolabs) per ml to the AST media for 96 hours.

TRANSFECTION OF MOUSE ASTS WITH SIRNA OLIGOS

Transfection of mouse post-mitotic ASTs with siRNAs was crucial for the investigation of the effect that the transient knockdown of the cohesion-associated proteins have in G_1/G_0 phase of the cell cycle. SiRNAs are cut from shRNA or from long synthetic dsRNA by the Dicer enzyme, and are later separated into two short strands, one of which binds to the target mRNA and cleaves it, preventing the target protein from being made. In the siRNA approach, specific siRNAs are synthesised to silence specific proteins in target cells. Because siRNA is composed of RNA and is inherently fragile, it is difficult to get into the cytoplasm and, once there, is quickly degraded. This means that relatively high doses (i.e. in the nmol range) are needed to achieve the desired level of gene silencing (Moore et al., 2010).

For this purpose, NSCs were initially platted with respect to optimal differentiation conditions and after 24 hours, they were differentiated into ASTs in the presence of BMP4. After an additional 24 hours, *Rad21* was depleted by infections with 10^7 plaque-

forming unit (PFU) of Adenovirus-CMV-Cre (Cre-Ad, Vector Biolabs) per ml to the AST media for 48 hours. After 48 hours, ASTs *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ were re-plated at optimal confluence (60%; 400,000 cells per well in 6-well plates) and after an additional 24 hours, ASTs were transfected according to RNAi Max Lipofectamine protocol. Each culture (each well per 6-well plate) was transfected with 50 µM of the mouse siGENOME pool, 4.2 µl of RNAiMax Lipofectamine in 200 µl of OPTIMEM (GIMCO) and a total volume of 1.7 ml per well. The transfections were performed with the respective siGENOME pools for Stag1 (M-041989-01), Stag2 (M-057033-01), CTCF (M-044693-01), Esco1 (M-066117-00) and siGENOME pool non-targeting control (D-001206-13) (Dharmacon). At specific timepoints, between 48 hours and 72 hours post-transfection, ASTs were collected by using accutase (PAA) for 15 minutes at 37^{o}C followed by a wash with media and centrifugation at 1,200 rpm for 4 minutes. The supernatant was poured off and cell pellets were collected same as above for downstream experiments.

Prior to the transfection experiments, the efficiency of the transfection in both NSCs and ASTs was evaluated by performing the transfections as described above using the mouse siGLO Green Transfection Indicator (siGLO-488, Dharmacon) and 48 hours post-transfections performing FACS analysis (described in CELL BIOLOGY METHODS).

5.2. CELL BIOLOGY METHODS

CELL CYCLE ANALYSIS AND FLOW CYTOMETRY OF MOUSE ASTS

Propidium iodide (PI) is a DNA binding dye, which is used to quantify the fluorescence intensity of the stained cells at certain wavelengths and correlates with the amount of DNA they contain. Thus, PI is very useful for cell cycle analysis since during the cell cycle the DNA content of cells changes were being duplicated in G_2/M phase of the cell cycle from what it was in G_1.

For this purpose, NSCs were collected at different time-points after BMP4 addition and fixed in 70% ethanol (10^6 of cells in 1 ml of 70% EtOH). The fixed cells were stored at 4^{o}C and then when all time-points were collected, they were stained with PI solution (100 µg/ml PI, 100 µg/ml RNase A and 0.05% Triton X-100 in PBS) for 40 minutes on ice at a concentration of $1x10^6$ cells per ml. Cells were monitored for clumping and passed over a 70 mm filter before proceeding to flow cytometry on a CyAn ADP Analyzer (Beckman Coulter) (10,000 events count per sample). Viable cells were selected using scatter plots of forward scatter (FSC) against site scatter (SSC) of each sample indicating size and the granularity of the cell, respectively. PI has a maximum emission of 605 nm so PI staining in this experiment was measured accordingly. The results were represented using the Flo-Jo software.

FLOW CYTOMETRY AFTER SIRNA TRANSFECTION OF MOUSE ASTS

NSCs and ASTs were collected as described above at 48 hours after siGLO-488 transfection, the supernatant was poured off and cell pellets were re-suspended as $10x10^6$ per ml of sorting media in order to be sorted for green-positive cells using the 488 nm laser (10,000 events count per sample). Viable cells were selected using scatter plots of FSC against site scatter SSC of each sample indicating size and the granularity of the cell, respectively. Then, GFP positive cells were sorted using the 488 nm blue laser (Log GFP Intensity > 10^3). The results of the cell sorting experiments were represented using the Flo-Jo software.

5.3. MOLECULAR BIOLOGY METHODS

PROTEIN EXTRACTION AND FRACTIONATION METHODS

Whole cell lysates

Whole protein lysate was purified for protein analysis from NSC and AST pellets. Cell pellets were collected and re-suspended in RIPA buffer (30 µl RIPA buffer per $5x10^6$ cells) and 25x protease inhibitor mix EDTA-free (Invitrogen) following incubation on ice for 10 minutes. Then, cell lysates were centrifuged at 13,000 rpm for 10 minutes at 4^oC and protein containing supernatants were stored at -80^oC.

Protein fractionation

Soluble and chromatin-bound protein fractions were collected separately to investigate the effect of the knock-down of Rad21 and the cohesin-associated proteins, on chromatin-bound proteins of interest. Initially, cell pellets were re-suspended in Fractionation Buffer 1 (10 mM Hepes pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M Glycerol, 1 mM DTT, 0.1% TritonX-100; 10^6 cells in 40 µl) and incubated on ice for 10 minutes. The soluble fraction was collected as the supernatant after spinning at 1,300 rcf for 10 minutes at 4^0C and the pellet was re-suspended in Fractionation Buffer 2 (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT; 10^6 cells in 40 µl) and incubated on ice for 30 minutes. The nuclear soluble fraction was collected as the supernatant after spinning at 1,900 rcf for 10 minutes at 4^oC and the pellet representing the chromatin fraction was re-suspended in 2x Laemmli Buffer (Bio-rad) (10^6 cells in 40 µl).

Protein salt-dependent fractionation (Martini extracts)

Soluble and chromatin-bound protein fractions were collected separately and a certain concentration of NaCl was used to remove salt-extractable from the chromatin-bound proteins in order to understand the effect of the *Rad21* depletion on chromatin-bound proteins of interest according to the protocol described in (Cook et al., 2011; Martini et al., 1998). Initially, the media was removed carefully from the culture plates and 10 ml of Buffer E (20 mM Hepes-KOH pH7.8; 5 mM KOAc, 0.5 mM $MgCl_2$, 0.5 mM DTT) per 150 cm^2 dish was added slowly and cells were incubated for 10 minutes at 4^oC. Then, Buffer E was removed, 50 ml of Buffer E with 25x protease inhibitor mix EDTA-free

(Invitrogen) per plate was added and cells were incubated for 10 minutes at 4°C. Buffers were removed and after plates were dried for 2 minutes, cells were removed with a cell scraper carefully so that the nuclei remained intact. Cells were transferred in a loose homogenizer vessel where the cells were mechanically lysed with 25 strokes. The soluble fraction was collected with the supernatant after spinning at spin 1,500 rcf for 5 minutes at 4°C and Buffer N (10% glycerol and 1 M NaCl topped up to 10 ml with Buffer E) was added to the pellet respectively so that final salt concentration was 300 or 600 mM NaCl for the experiments shown here. Then, nuclear pellets were incubated for 90 minutes in the rotator at 4°C. The volume of Buffer N that was added to the nuclear pellets was calculated as follows: V2=V1/0.67 (Final salt concentration 600 mM NaCl), where V1=Volume of the Pellet (salt-extract/chromatin bound), V2=Volume of Buffer N added to the pellet. To collect the salt-extractable chromatin fraction, pellets were span at 14,000 rcf for 20 minutes at 4°C and supernatant was removed. The pellets represented the salt-resistant chromatin fraction and were re-suspended in 2x Laemmli Buffer (Bio-rad). Protein concentration was calculated in the soluble fraction samples using the Bradford assay (described in PROTEIN QUATIFICATION AND WESTERN BLOT) and the volume used from each fraction was the same percentage of the total volume used in the respective sample of soluble, salt-extractable and insoluble fraction. Between samples, 10 µg of the soluble fraction of each sample were loaded per well of the SDS-PAGE gel so that the downstream Western blot analysis was performed.

PROTEIN QUANTIFICATION AND WESTERN BLOT

Protein samples were quantified by using the Bradford assay (BioRad) and the BSA standard curve method. Western blot analysis was performed for Rad21 (Abcam, ab992), Smc3 (Abcam, ab9263), Cre Recombinase (CreR) (Novagen, 69050), CTCF (Abcam, ab70303), Histone 3 (Abcam, ab1791), Hsp90 (Abcam, ab13494), tubulin (alpha) (Sigma, T5168), Stag2 (Bethyl, A300-158A), Stag1, Pds5a and Pds5b antibodies from A. Losada and Smc3 acetylation from K. Shirahige laboratories. Each protein sample (10µg) was diluted in a total volume of 15-30 µl with 2x Laemmli Buffer (Bio-rad). The samples were boiled at 95°C for 5 minutes and loaded on 6-8% SDS-PAGE gels or 4-12% SDS-PAGE pre-cast gels (Bio-Rad) along with 10 ml of protein ladder (Bio-Rad). Gels were electrophorised at 100 Volts in SDS-PAGE 1x Tris-Glycine Running buffer (1x Tris-Glycine with 0.1% SDS) and proteins were transferred onto an

immune-Blot nitrocellulose membrane (Bio-Rad) by wet transfer at 160 mA per 1.0 mm gel for 2 hours in a wet transfer apparatus (Bio-Rad) with an appropriate amount of SDS-PAGE 1x Tris-Glycine transfer buffer (1x Tris-Glycine with 20% methanol). After blocking [5% (w/v) milk (Sigma), 0.1% Tween20 (Sigma) in Tris Buffered Saline (TBS)] for 1 hour at room temperature, incubation with primary and secondary antibodies was performed overnight at 4^{o}C and 1 hour at room temperature in blocking solution, respectively. Before and after the incubation of the secondary antibody three washes of 10 minutes with TBS-Tween (0.05% Tween20 in TBS) were performed at room temperature. Finally, the detection was performed using ECL plus kit (GE healthcare) and they were developed according to the manufacturer's protocol.

IMMUNOFLUORESCENCE AND MICROSCOPY

ESCs, NSCs and ASTs were plated in 6-well plates (300,000 cells per well) on glass coverslips laminin-coated (10 ng/ml in PBS; overnight). Initially, cells were fixed with 4% paraformaldehyde for 10 minutes, washed with PBS, and permeabilised with PBS-Triton X-100 (0.1%) for 10 minutes. For specific experiments, to visualise chromatin bound proteins only, a pre-extraction step was added before the fixation step with PE Buffer (100 mM PIPES, 30% glycerol, 1 mM EGTA, 5 mM $MgCl_2$, and 0.1% Triton X-100) for 5 minutes at room temperature. Then, fixed cell cultures were blocked with 3% goat serum in PBS-Triton X-100 (0.1%) for 30 minutes. Incubation with both primary and secondary antibodies was performed for 30 minutes at room temperature in blocking solution. Three washes with PBS-Triton X-100 (0.1%) were performed between antibody incubations and after incubation with the secondary antibody. Finally, coverslips were washed with PBS and mounted with Vectashield containing 1.5 mg/ml DAPI. Images were acquired using a Leica TCS SP5 confocal microscope at a 63x magnification. Stacks were collected with a z step size of 0.3 mm. Nuclear volume and staining intensity were quantified by 'Find Objects' feature of the Volocity software (Perkin Elmer) and the box or violin plots were generated using R programing language.

DNA ANALYSIS

Genomic DNA purification

Genomic DNA was extracted from NSCs and ASTs using the QIAamp DNA Mini Kit (Quiagen) according to the DNA purification from blood or body fluids protocol. The DNA concentrations of the samples were then measured using the NanoDrop Ultraviolet (UV) spectrophotometer.

Quantitative Polymerase Chain Reaction (qPCR)

Genomic DNA was used as a template for the quantitative Polymerase Chain Reaction (qPCR) in order to determine the percentage of the depletion at the genomic locus between exons 5 and 6 of *Rad21* in *Rad21*$^{\Delta/\Delta}$ NSCs and ASTs. The final volume of genomic DNA was diluted up to a concentration of 100 ng/ml and DNA was used as a template for qPCR. The reaction was performed according to the LC480 SYBR Green 1 Master protocol (Roche) by using 2 µl of the diluted DNA library, 1.5 µl of each primer (10 mM), 10 µl SYBR green Master Mix (Roche) made up with water to a volume of 20 µl per reaction. The qPCR conditions used were as follows: 95°C for 5 minutes (1 cycle); 95°C for 10 seconds; 60°C for 30 seconds; 72°C for 30 seconds and single read (40 cycles). The qPCR was performed using the LightCycler 480 System (Roche). Amplification results were obtained using the manufacturer algorithm for 'fit points'. The quantification of the gene copy number under investigation (i.e. *Rad21*) was relative to the expression of the housekeeping genes (*Gapdh*) (primer sequences in **Table 1**).

RNA ANALYSIS

RNA extraction and cDNA synthesis

Total RNA was extracted from ESC, NSC and AST pellets using the RNeasy Kit (Qiagen) and was eluted as 10x10^6 cells in 50 µl of RNase/DNase-free H_2O. As a template for cDNA synthesis, total RNA of 1 µg in 11 µl of RNAase-free water was used per reaction. The cDNA reaction was performed according to the SuperScript II Reverse Transcriptase (1 µl/reaction, 200 units) protocol (Invitrogen) using 50-250 ng of random primers (RP) (5 µM, Invitrogen), 4 µl of 5x First Strand Buffer (Invitrogen), 10 mM DTT (0.1 M, Invitrogen), 0.5 mM dNTPs (10 mM, Yorkshire Biosciences), 40 Units

of RNase OUT Ribonuclease Inhibitor (Invitrogen) and 200 Units of SuperScript II reverse transcriptase (Invitrogen) made up with water to a total volume of 20 µl per reaction. cDNA synthesis was performed on a Mastercycler Realplex PCR machine (Eppendorf) with the following conditions: 65^{o}C for 5 minutes; 25^{o}C for 5 minutes; 50^{o}C for 60 minutes; 70^{o}C for 15 minutes.

qRT-PCR

The final volume of cDNA was diluted up to 250 µl of dH_2O to get to an approximate concentration of 100 ng/µl. cDNA was used as a template for qPCR. The qPCR reaction was performed according to the LC480 SYBR Green 1 Master protocol (Roche) by using 200 ng in 2 µl of the diluted cDNA library, 0.75 µM primer (10 µM, Sigma), 10 µl of 2x SYBR green Master Mix (Roche) made up with water to a volume of 20 µl per reaction. The qPCR conditions used were as follows: 95°C for 5 minutes (1 cycle); 95°C for 10 seconds, 60°C for 30 seconds, 72°C for 30 seconds and single read measurement, (40 cycles). The qPCR was performed using the LightCycler 480 System (Roche). Amplification results were obtained using the manufacturer algorithm for fit points. The quantification of the expression of the gene under investigation (i.e. *Rad21*) was relative to the expression of the housekeeping genes (i.e. *Tbp, Ubc*) (primer sequences in **Table 1** in the APPENDIX).

RNA extraction, RNA-Seq libraries and analysis

Total RNA was extracted from pellets from 5x10^{6} *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ NSCs and ASTs in duplicates using the mRNeasy Kit (Qiagen) according to the manufacturer's protocol. Then, samples were treated with DNase (Ambion) according to manufacturer's instructions. Quality assessment of the RNA samples was performed using the RNA 6000 Nano Kit in the 2100 Bioanalyzer (Agilent). An RNA integrity number (RIN) of at least 8 was obtained from the samples which were used for library preparation. RIN is a tool, which allows the evaluation of RNA integrity by considering the entire electrophoretic trace of the RNA sample. This includes the presence or absence of degradation products. In this way, interpretation of an electropherogram is facilitated, comparison of samples is enabled and repeatability of experiments is

ensured. The assigned RIN is independent of sample concentration, instrument and analyst therefore becoming a de facto standard for RNA integrity.

RNA samples were sent to Illumina Inc. for the RNA-seq library preparations. Strand-specific polyA+ libraries were prepared and sequencing was performed using the Illumina platforms (Genome Analyzer IIx) (library information in **Table 4** in the APPENDIX). Filtered reads were aligned to the genome using TopHat and transcript assembly was performed using Cufflinks (Trapnell et al., 2010) and the aligned reads were imported in the UCSC Genome Browser to be visualised with respect to the mouse genome. Expression was expressed as fragments per kilobase of exon per million mapped reads (FPKM). The de-regulated genes were defined the group of genes that had z-score or standard score more than or equal to 1.5 where z-score allowed the calculation of the probability of an FPKM occurring within a normal distribution and enables the comparison two FPKM values that are from different normal distributions.

Finally, the results presented here were statistically analysed and graphically represented in by using R statistics software. The gene enrichment analysis regarding the pathways that the de-regulated genes were involved in was done by Protein ANalysis THrough Evolutionary Relationships (Panther) classification system.

CHROMATIN IMMUNOPRECIPITATION (CHIP)-SEQ LIBRARIES AND ANALYSIS

Genome-wide mapping of protein–DNA interactions and epigenetic marks is fundamental for the deeper understanding of transcriptional regulation. A precise map of binding sites for transcription factors and other DNA-binding proteins is vital for reading the gene regulatory networks that underlie various biological or disease-related processes. So Chromatin Immunoprecipitation experiments followed by sequencing (ChIP-seq) allow for the detection protein–DNA binding events in high-resolution and throughout the genome (Furey, 2012).

The main steps of the protocol are the following (**Figure 34**): a) crosslinking proteins with DNA using formaldahyde solution; b) Chromatin fragmentation by sonication; c) Immunoprecipitation for the protein of interest; d) DNA purification, library preparation and sequencing by using the Illumina standard protocol. Challenges in the standard ChIP–seq protocol have motivated recent advancements in this approach; such as

modifying the protocol so that the resolution is increased (Rhee and Pugh, 2012) or challenges in the analysis of the data such as the inability to quantitatively compare datasets genome-wide, which has been proposed to be overcome by introducing equal amounts of reference genomes into the ChIP so that in the analysis step the original ChIP is normalised to the exogenous sample (Orlando et al. 2014).

Preparation of chromatin and fragmentation

Cell pellets from NSCs and ASTs were collected and re-suspend in neural stem cell media with 1% formadehyde (SIGMA) (35-40x10^6 cells in 20 ml media) for 20 minutes at room temperature (**Figure 34**). Crosslinking reaction was terminated by the addition of 0.125 M glycine for 5 minutes at room temperature. Then, cell pellets were collected by spinning at 2,500 rpm for 5 minutes at 4^{o}C and washed twice with PBS (GIBCO) (10 ml per pellet) before they were stored at -80^{o}C not more than three weeks. Cell pellets were thawed at 37^{o}C briefly and washed with Wash Buffer 1 (10 mM Hepes pH 7.5, 10 mM EDTA, 0.5 mM EGTA, 0.75% Triton X-100) and Wash Buffer 2 (10 M Hepes pH 7.5, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) for 10 minutes at 4^{o}C, respectively. Then, nuclei were collected by spinning at 2,500 rpm for 5 minutes at 4^{o}C and lysed with Lysis Buffer without SDS for 30 minutes (150 mM NaCl, 25 mM Tris-HCl pH7.5, 5 mM EDTA, 1% Triton X-100, 0.5% NaDeoxycholate) at a final concentration of 2x10^6 cells in 100 µl. Then, a final concentration of 0.2% SDS was added and NSC or AST lysates were sonicated using the Bioruptor Pico Sonication system (Diagenode) for a total of 4 or 8 minutes at 4^{o}C, respectively. Sonicated lysates were span at 13,200 rpm for 30 minutes at 4^{o}C so that non-soluble chromatin was removed and supernatants were used for the following immunoprecipitation step.

Quantity and quality control of sonicated chromatin

Quantity and quality controls were performed at this stage by a sample representing 10% of the initial sonicated lysates (**Figure 34**). Initially, proteins were decross-linked from DNA by adding to 10 µl of the sample, 10 µg/ml of proteinase K (Bioline) and 10 mM of Tris-HCl (pH7.5) up to a total volume of 100 µl, overnight at 65^{o}C. Then, DNA was extracted by the addition of equal volume of phenol/chloroform and the top aqueous phase was collected via spinning at 12,400 rpm for 5 minutes at room

temperature (herein refered as phenol/chloroform extraction method). Finally, DNA was precipitated by the addition of 0.3 M sodium acetate (SIGMA), 1 µl of GlycoBlue and 2.5 volumes of 100% EtOH and incubation at -80^{o}C for minimum 2 hours (herein reffered as EtOH precipitation). Then, precipitated DNA was isolated by spinning at 13,200 rpm for 30 minutes at 4^{o}C, washed twice with 70% EtOH, air-dried for 30 minutes and re-suspend in 30 µl of 1x Tris-EDTA (TE). For quantity control, the Qubit quantitation system with the dsDNA HS kit (Invitrogen) was used to quantify the DNA content of the sample according to the manufacturer's instructions and approximate the total amount in the original sample (about 6x10^{6} ASTs yield about 25 µg of DNA). For quality control of the sonication efficiency and the extent of DNA fragmentation, 15 µl of the purified DNA were electrophoresed on a 1% agarose gel to see the extent of sonication. Optimal condition was considered a smear representing a wide range of fragment sizes with the majority of fragments being between 2-6 kb. At this step chromatin was frozen at -80^{0}C before use and protease inhibitors were added upon thawing.

Chromatin immunoprecipitation (ChIP)

For the preparation of a ChIP-Seq library 100 µg of chromatin was an optimal amount as an input to the chromatin immunoprecipitation step (ChIP). For the Rad21 ChIP, 40 µg of the Rad21 antibody (1 µg/µl, Abcam Ab992) conjugated with 100 µl Protein G (PG) Dynabeads (10 mg/ml, Life Technologies). In addition, a control for unspecific chromatin binding was also included for every ChIP experiment, where 4 µg Rabbit anti-mouse polyclonal immunoglobulins (10 mg/ml, Dako) control was conjugated with 20 µl PG beads. Initially, PG beads were washed with 1 ml of Wash Buffer A (50 mM Tris-HCl pH8.0, 150 mM NaCl, 0.1% SDS, 0.5% NaDeoxycholate, 1% NP-40, 1 mM EDTA) twice on the magnetic rack and then, re-suspended in Wash Buffer A to a final concentration of 100 ng/µl and were blocked with 0.1 mg/ml of BSA (NEB) for 1 hour at 4^{o}C, rotating. Then, PG beads were washed twice with 1 ml Wash Buffer A and re-suspended in Wash Buffer A to a final concentration of 100 ng/µl. Conjugation of the blocked PG beads with the antibody at the analogies described above was performed for 4 hours at 4^{o}C, rotating. At the same time, chromatin was diluted with Lysis Buffer at a final concentration of 0.1% SDS and pre-cleared with the same amount of PG beads as conjugated with the antibody at the same conditions. Finally, the PG beads with conjugated antibody were pulled together with the chromatin lysates and

incubated over-night at 4^{o}C, rotating. Before the conjugation, a percentage of 1% of the input sample was stored at -80^{o}C. After 16-20 hours, lysates were removed from beads and kept at -80^{o}C as the post-IP samples and DNA was eluted with consecutive washes with 1 ml of Wash Buffer A, B (50 mM Tris-HCl pH8.0, 500 mM NaCl, 0.1% SDS, 0.5% NaDeoxycholate, 1% NP-40, 1 mM EDTA) and C (50 mM Tris-HCl pH8.0, 250 mM LiCl, 0.5% NaDeoxycholate, 1% NP-40, 1 mM EDTA) for 10 minutes at 4^{o}C, rotating. Finally, a quick wash with 1x TE buffer was performed and DNA on the beads was decrosslinked with 400 µl of Elution Buffer (1% SDS, 0.1 M $NaHCO_3$, and 75 µg/ml Proteinase K up to a volume of 400 µl with dH_2O) at 65^{o}C overnight. Then, samples were treated with 37.5 µg/ml of RNase A (Invitrogen) for 2 hours at 37^{o}C and DNA was precipitated as described in quality and quantity control paragraph.

Western blot and ChIP-qPCR quality control

To check whether the fixation conditions and the sonication conditions were optimal in order to precipitate most of the protein of interest, IP sample, an input sample and a sample from the post-IP supernatant were electrophorised on an SDS-PAGE gel and the amounts of protein in those fractions were compared. For such tests, IPs of 20 µg of chromatin were used. For this protocol, after the washes with Wash Buffers A/B/C and TE, the beads were re-suspended in a volume of Lysis Buffer (no SDS) equal to the original volume of the chromatin you used for the IP. Samples were loaded on a 1.5 mm SDS-PAGE gel and the Western blotting protocol was followed.

Additionally, an IP sample, an input sample and a sample from an IgG control were used as an input to the qPCR. For such tests, IPs of 20 µg chromatin were used and 5 primer sets in replicates were tested. Primers were previously described in (Parelho et al., 2008) and found to be positive for Rad21 binding (i.e. Chr 9:4457) or negative (i.e. Xist, Chr 10:2448) (primer sequences in **Table 2** in the APPENDIX) (**Figure 34**).

ChIP-seq library preparation, sequencing and analysis

The optimal input DNA for the library preparation used was 50 ng using the NEBNext ChIP-Seq Library preparation kit for Illumina (NEB) according to the manufacturer's instructions. After each enzymatic reaction, the DNA was extracted using 1.8 volumes of Agencourt AMPureXP beads (Beckman Coulter) and eluted into the volume required

for the next reaction. The reactions were the following: 1) Repair the ends: 50 ng of DNA, 5 µl of 10x End Repair Buffer and 1 µl of End Repair Enzyme up to a total volume of 50 µl of water and incubated for 30 minutes at 20^{o}C; 2) Add A-tails: 44 µl of the DNA from previous reaction, 5 µl of 10x dA-tailing Buffer and 1 µl of the Klenow fragment (3'-5' exo) up to a total volume of 50 µl of water and incubated for 30 minutes at 37^{o}C; 3) Ligate the Illumina adaptors: 19 µl of the DNA from previous reaction, 6 µl of the 5x Quick Ligation Reaction Buffer, 1 µl of the Illumina Adaptors (1.5 µM) and 4 µl of the Quick T4 DNA Ligase up to a total volume of 30 µl and incubated for 15 minutes at 20^{o}C. After the incubation 3 µl of the USER Enzyme Mix was added to the reaction and samples were incubated for 15 minutes at 37^{o}C. At this point DNA was eluted into 23 µl dH_2O following the AMPure beads purification; 4) PCR: 11.5 µl of DNA, 12.5 µl of the Phusion Polymerase Mix, 0.5 µl of the Universal PCR Primer and 0.5 µl of the Index Primer up to a total volume of 25 µl and incubated as follows: 98^{o}C for 30 seconds (1 cycle); 98^{o}C for 10 seconds, 65^{o}C for 30 seconds, 72^{o}C for 30 seconds (12-15 cycles); 72^{o}C for 5 minutes (1 cycle). If two libraries were multiplexed later prior to sequencing, Index 6 and Index 12 from the NEBNext ChIP-Seq Library preparation kit for Illumina (NEB) were used. The correct combinations, which depend on the number of libraries multiplexed, was determined according to NEB/Illumina manufacturer's protocol.

Following the PCR, samples were purified using a 1:1 ratio of AMPure beads. This removed the majority of adaptors and samples were eluted in 6 µl of 1x TE and 1 µl was used for the Bioanalyzer according to the DNA 1000 kit (Agilent). Optimal ChIP library Bioanalyzer smear showed the average size of the library at about 300 bp and the consequent sequencing of the libraries was performed following the Illumina protocol for the MiSeq platform. Based on the above Bioanalyzer results, the libraries were multiplexed. Here, the molarity was about 10 nM as predicted by the Bioanalyzer and the library mix was diluted to a final concentration of 2 nM with water in a volume of 10 µl. Then, equal volume of 0.2 N NaOH was added and samples were incubated for 5 minutes at room temperature. Finally, 980 µl Hybridisation buffer (Illumina) was added to the sample on ice and 500 µl of the sample were further diluted with 500 µl of Hybridisation buffer up to 1 ml total volume to achieve 10 pM as final concentration of the sample. For sequencing on the MiSeq platform, 600 µl of the sample above are optimal and a single-end read length of 75 bp sequencing to collect ChIP-Seq data.

The analysis of the ChIP-Seq dataset was performed as described in (Sofueva et al., 2013) by Dr C. Barrington, where the coverage of mapped reads was assigned to 50

bp bins across the genome. Quantile value for each bin was transformed by log10(1-quantile(v)) to compare between datasets. Bins with signal values above 2.3 were identified as Rad21-bound.

CHROMOSOME CONFORMATION CAPTURE (3C) METHODS

Chromosome conformation capture (3C) methods have been developed to study genome organisation in higher resolution by measuring quantitatively chromatin interactions (**Figure 3**). Briefly, 3C techniques include the following main steps: a) crosslinking proteins with DNA using folmadehyde solution so that DNA fragments are crosslinked together when they are in close proximity to each other in three-dimensional space within intact nuclei; b) DNA fragments are digested with a restriction endonuclease and ligated and the closer the fragments are, the more often they will ligate to each other; c) PCR of the ligation junctions. In recent years, 3C has been modified to allow for simultaneous assessment of interactions between multiple loci (4C (de Wit and de Laat, 2012) and 5C (Dostie et al., 2006)), as well as the discovery of novel interactions in the presence (ChIA-PET) (Fullwood et al., 2009a) or absence (Hi-C) (Lieberman-Aiden et al., 2009) of a protein-based pre-enrichment step (described in CHAPTER 1).

Here, the candidate-based high-resolution 4C-Seq (van de Werken et al., 2012) and the genome-wide Hi-C protocols were used, which both follow the above principle and were performed *in nucleo*, which preserved the natural environment of these interactions (**Figure 26**). The aim of 4C-Seq is to measure how often a particular restriction fragment (herein referred as bait) interacts with many other restriction fragments in the genome allowing the modeling of the three-dimensional chromatin folding from the perspective of a specific locus of interest. More specifically, following the first ligation step DNA concatamers are formed, containing multiple restriction fragments, which precludes efficient amplification and sequencing across the region of interest. Therefore, the second round of digestion and ligation was used to reduce the size of the circularised fragments. Importantly, using two rounds of 4 bp restriction enzymes significantly increases the complexity of the fragment pool and thus the resolution of the resulting data. Finally, libraries were prepared by amplifying the circularised products after the second round of digestion and ligation using primers specific to the site of interest and containing an overhang with the Illumina adaptor sequences (**Figure 26**). By comparison the aim of the HiC protocol is to measure how

frequently all fragments in the genome interact with all fragments (all-to-all approach) and therefore probing for global genome architecture. This is achieved by the incorporation of a biotin-labelled nucleotide at the ligation junction, enabling selective purification of chimeric DNA ligation junctions followed by deep sequencing (**Figure 26**).

Cell lysis and chromatin preparation for 3C methods (4C-Seq and Hi-C)

Cell pellets were isolated and re-suspend in neural stem cell media with 1% formaldehyde (SIGMA) ($10x10^6$ cells in 20 ml media) for 20 minutes for NSCs and 30 minutes for ASTs at room temperature. Crosslinking reaction was terminated by the addition of 0.125 M glycine for 5 minutes at room temperature. Then, cell pellets were isolated by spinning at 2500 rpm 5 minutes at 4^{o}C and washed twice with PBS (10 ml per pellet, GIBCO) before they were stored at -80^{o}C. Cell pellets were thawed and lysed in 10 ml of Lysis Buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% NP-40) for 20 minutes on ice. Then, nuclei were collected by spinning at 1800 rpm for 5 minutes at 4^{o}C and re-suspended in 0.5 ml 1.2x DpnII buffer and transferred to a Protein LoBind tube (Eppendorf). At this point the concentration of nuclei was quantified using a haemocytometer with 2.5 µl of the sample and 47.5 µl of Crystal violet dye (1% final dilution).

CIRCULARISED CHROMOSOME CONFORMATION CAPTURE (4C-SEQ)

Digestion (DpnII) and Ligation I, Digestion II (Csp6I) and Ligation II

An initial number of $7\text{-}8x10^6$ nuclei were re-suspended in 0.5 ml of 1x DpnII buffer and 0.3% SDS and aliquots were incubated at 37^{o}C for 1 hour. Then, 2% Triton X-100 was added per aliquot and aliquots were incubated at 37^{o}C for 1 hour. At this point, 10 µl of the aliquots were collected and stored as the undigested control sample. Then, 750 Units of DpnII were added to each aliquot and aliquots were incubated at 37^{o}C overnight, rotating. At this point, 10 µl of the aliquots were also collected as a digested for 24 hours digestion control and DpnII buffer was re-newed (1.2x DpnII Buffer with 0.3% SDS and 2% Triton X-100) with 750 Units of DpnII and aliquots were incubated at 37^{o}C overnight, rotating. Finally, 10 µl of the aliquots were collected as a digested for 2 days control. All control samples were decrosslinked and 30 µl per sample of the aqueous phase after phenol/chloroform extraction was run on a 1% agarose gel. A

second digestion should lead to a shift in the average fragment size in the case of NSCs and ASTs. Different cell types can vary and occasionally one night of digestion may be enough. Generally, a pool of fragments below the 3 kb mark is optimal as an input to the next step of the protocol.

Digested fragments were collected by spinning briefly, re-suspended in 100 µl of the 1x T4 DNA Ligase buffer with 1600 Units of T4 DNA Ligase and incubated at 16^{o}C overnight. At this point, ligation efficiency was evaluated the same way digestion was checked by 2.5 µl of the sample. A successfully ligated sample usually runs as a single band, slightly smaller in size than the undigested band (**Figure 26**). However, note that a different profile for ligated samples has been reported in (van de Werken et al., 2012) depending on the cell type. Based on an optimal ligation profile, 300 µl of 1x TE (pH 8.0) and 1 mg/ml of Proteinase K (Bioline) were added and aliquots were incubated overnight at 65^{o}C to reverse crosslinking reaction. Then, 75 µg/ml of RNase A (Invitrogen) were added and aliquots were incubated at 37^{o}C for 2 hours. Then, DNA was extracted by phenol/chloroform extraction method, was precipitated using EtOH and re-suspend in 50 µl 1x TE.

The concentration of purified ligated DNA was quantified using the Qubit quantitation system with the dsDNA HS kit (Invitrogen) and 1 µl was collected as a sample of undigested controls II. Then, the second digestion reaction was set up as follows: about 19 µl (5.5 ug) of sample, 30 µl of 10x Buffer B (NEB) and 120 Units of Csp6I (NEB) up to a total volume of 300 µl with water. The reactions were incubated overnight at 37^{o}C. At this point, a sample of 10 µl was collected to check the digestion efficiency on a 1% agarose gel. Similarly to DpnII digestion, a pool of fragments below the 3 kb mark is optimal as an input to the next step of the protocol (**Figure 26**). Based on optimal digestion profiles, DNA was purified via phenol/chloroform extraction method and precipitated using EtOH and re-suspended in 50 µl of water. Ligation reaction was set up as follows: 50 µl of the sample, 600 µl of the 10x T4 DNA Ligase buffer, 1,600 Units of T4 DNA Ligase up to a total volume of 6 ml with water and samples were incubated overnight at 16^{o}C. Finally, DNA was purified by using the phenol/chloroform method, precipitated using EtOH and re-suspended in 30 µl of Tris-HCl (10 mM, pH 7.5).

4C-Seq library preparation, sequencing and analysis

DNA concentration was quantified using the Invitrogen Qubit quantitation system and primer mix was prepared (0.85 µg/µl in 10 mM Tris pH 7.5). Next the PCR reaction used primers from the perspective of any bait sequence. Once the site is chosen, primers are designed to the two ends of the nearest DpnII-Csp6I restriction fragment. The primer at the DpnII end should contain the DpnII recognition site. The primer at the Csp6I end can be away from the recognition site, usually within a restricted distance (e.g. 100 bp). The database of pre-designed 4C-Seq primers for mouse can be found in (van de Werken et al., 2012) (primer sequences in **Table 3** in the APPENDIX). PCR reaction was performed with the Expand Long Template PCR System (Roche) and was set up as follows: 100 ng of 4C template, 2.5 µl of 10x Buffer 1, 0.2 mM dNTPs (Invitrogen), 0.05 µg/µl of the primer mix, 0.35 µl of the Expand Long Template Polymerase (Roche) up to a total volume of 25 µl of water. Reaction conditions were the following: 94^{o}C for 3 minutes (1 cycle); 94^{o}C for 10 seconds, 55^{o}C for 1 minutes, 68^{o}C for 3 minutes (30 cycles); 68^{o}C for 10 minutes (1 cycle). Then, the PCR product was purified with the High Pure PCR Product Purification kit (Roche) according to the manufacturer's instructions to remove primer dimers and concentrated to about 15 µl using AMPureXP beads according to the manufacturer's instructions. Finally, 1 µl of the sample was evaluated on a Bioanalyzer chip using the Agilent DNA 1000 kit. In optimal conditions, the Bioanalyzer profile of each library should contain two prominent bands (one for the undigested product and one for the self-circularised product) on a smeared background of fragments (**Figure 26**). If the two theoretically prominent bands are too large or too small, they will not be visible on the profile. This variable product size does make the sequencing reaction less efficient; however, usually more than 95% of clusters pass the filters.

Based on the above Bioanalyzer results, each library was diluted to 10 nM and equal volumes of the libraries intended to be sequences in the same MiSeq flow cell were mixed. The library mix was quantified using the Invitrogen Qubit quantitation system and the theoretical molarity of the library mix was estimated using the following formula: [Library mix [c] µg/µl : (650 x 600)] x 10^{9} = nM library mix]. Here, the molarity was about 10 nM as predicted by the Bioanalyzer and the library mix was diluted to a final concentration of 2 nM with water in a total volume of 10 µl. Then, samples were prepared for sequencing in the MiSeq platform (described in ChIP-Seq method). For sequencing, a single-end read length of 50 bp or more sequencing is optimal to collect 4C-Seq data. A detailed description of the pipeline used to visualise data from 4C-Seq

experiments are described in (van de Werken et al., 2012). Briefly, libraries were classified based on the sequencing reads containing the forward primer (Primer 1), which serves as a tag for individual baits. The first 18-20 bp of the reads was trimmed and the remaining 30 or more bp were mapped to the mouse genome. The pipeline visualises normalised reads by representing the median of the reads either as a black line using a sliding window of 5 kb or as a heatmap using a range of sliding windows from 2 to 50 kb.

HI-C

Digestion (HindIII), Fragment fill-in reaction and Ligation

An initial number of 7-8x10^6 nuclei were re-suspended in 0.5 ml of 1x NEB2 buffer and 0.3% SDS and aliquots were incubated at 37^{o}C for 1 hour. Then, 2% Triton X-100 was added per aliquot and aliquots were incubated at 37^{o}C for 1 hour. At this point, 10 µl of the aliquots were collected and stored as the undigested control sample. Then, 1500 Units of HindIII were added to each aliquot and aliquots were incubated at 37^{0}C overnight, rotating. At this point, 10 µl of the aliquots were also collected as a digested control. For ASTs, the digestion was performed with 0.6% SDS and 4% Triton X-100 and was efficient by 1 day of digestion. All control samples were de-crosslinked and 30 µl per sample of the aqueous phase after phenol/chloroform extraction was run on a 1% agarose gel.

Digested fragments were collected by spinning briefly, re-suspended in 100 µl of the 1x T4 DNA Ligase buffer with 1600 Units of T4 DNA Ligase and incubated at 16^{o}C overnight. At this point, ligation efficiency was evaluated the same way digestion was checked by 2.5 µl of the sample (**Figure 36**). After a quick-spin to isolate the nuclei, 103.4 µl of the mixture was added to the pellet [20 µl of 10x NEB2, 0.1 mM Biotin dNTP mix (0.36 mM of biotin-dCTP, dGTP, dTTP and dATP, Invitrogen), 25 Units of DNA polymerase I Large (NEB)] and an additional 85 µl of dH_2O. The samples were incubated at 37^{o}C for 45-50 minutes and they were inverted to mix every 15 minutes.

Digested fragments after fill-in reaction were collected by spinning briefly, re-suspended in 100 µl of the 1x T4 DNA Ligase buffer with 1,600 Units of T4 DNA Ligase and incubated at 16^{o}C overnight. At this point, ligation efficiency was evaluated the same way digestion was checked by 2.5µl. A successfully ligated sample usually runs

as a single band, slightly smaller in size than the undigested band (**Figure 36**). Based on an optimal ligation profile, 300 µl of 1x TE (pH 8.0) and 1 mg/ml of Proteinase K (Bioline) were added and aliquots were incubated overnight at 65^{o}C to reverse crosslinking reaction. Then, 75 µg/ml of RNase A (Invitrogen) were added and aliquots were incubated at 37^{o}C for 2 hours. Then, DNA was extracted by using the phenol/chloroform method, was precipitated using EtOH and re-suspend in 100 µl of 1x TE.

Hi-C library preparation, sequencing and analysis

DNA concentration was quantified using the Invitrogen Qubit quantitation system and to remove biotin from unligated ends, the following reaction was set up for 5 µg of DNA: 10 µl of 10x NEB2, 10 µl of 10x BSA, 5 Units of T4 DNA Polymerase 0.1 mM dATP, 0.1 mM dGTP topped up with dH_2O to a final volume of 100 µl. The reaction was incubated for 2 hours at 12^{o}C and 2 µl of 0.5 M EDTA (pH8.0) was added to stop the reaction. DNA (10x reactions, 40-50 µg of DNA) was phenol/chloroform extracted, EtOH precipitated and re-suspend in 100 µl 1xTE (10 mM Tris-HCl pH8.0; 0.1 mM EDTA). DNA concentration was quantified using the Invitrogen Qubit quantitation system and 4 µg of DNA were diluted up to 120 µl 1x TE to be sonicated by the S220 Covaris sonicator system (AFA Technology) for 140 seconds using the following settings: 20% duty cycle, 5 intensity, 200 cycle burst. Finally, 1 µl of the sample was evaluated on a Bioanalyzer chip using the Agilent DNA 1000 kit. In optimal conditions, the Bioanalyzer profile of each library should be between 150 to 250 bp sizes (**Figure 36**).

To repair the sonicated ends, the following reaction was set up for 4 µg of DNA: 16.8 µl 10x Ligase buffer, 0.25 mM dNTP mix, 15 Units of T4 DNA polymerase (NEB), 50 Units of T4 Polynucleotide Kinase (NEB) and 5 Units of DNA polymerase I Large (NEB) topped up with dH_2O up to a final volume of 168 µl. The reaction was incubated for 30 minutes at 22^{o}C and DNA was eluted in 30 µl 1x TE using MinElute kit (Quigen) according to the manufacturer's instructions. Size selection of 200-400 bp fragments was performed by using 1.5% agarose (UltraPure, Invitrogen) in 1x Tris-Boric Acid-EDTA (TBE) gels where its lane represented the 4 µg of DNA and each band was eluted in 18 µl 1x TE using Gel Extraction kit (Quigen) according to the manufacturer's instructions.

DNA concentration was quantified using the Invitrogen Qubit quantitation system and to select for the biotinylated ligated restriction sites, 180 µl of eluted DNA was incubated with 50 µl of Dynabeads MyOne Streptavidin C1 (Invitrogen) for 1 hour rotating at room temperature. Prior to the conjugation, the 50 µl Dynabeads were washed twice with 400 µl Tween buffer (5 mM Tris-HCl pH 8.0, 500 µM EDTA pH 8.0, 1 M NaCl, 0.05% Tween20) for 3 minutes rotating at room temperature and re-suspended in 180 µl of 2x no Tween buffer (5 mM Tris-HCl pH 8.0, 500 µM EDTA pH 8.0, 1M NaCl). After 1 hour incubation of the Dynabeads with the DNA, beads were washed twice with 400 µl of 1x no Tween buffer for 3 minutes rotating at room temperature and once with 400 µl of 1x NEB2.

Each aliquot of Dynabeads was re-suspended in 50 µl of 1x NEB2 and to add polyA-tails to the fragment ends, the following reaction was set up: 5 µl of 10x NEB2, 0.1 mM dATP, 5 Units of Klenow fragment (3'-5' exo) (NEB) topped up with dH_2O to a final volume of 100 µl. The reaction was incubated for 45 minutes at 37^{o}C and beads were washed twice with 100 µl of 1x no Tween buffer for 3 minutes rotating at room temperature and once with 100 µl of 1x T4 DNA ligase Buffer. Each aliquot of Dynabeads was re-suspended in 50 µl of 1x T4 DNA ligase Buffer and to ligate the Illumina adapters the following reaction was set up: 5 µl of 10x T4 DNA ligase Buffer, 1,200 Units of T4 DNA ligase, 6 pmol Pair-end (PE) Adapters 1.0/2.0 (Illumina) per µg of DNA (as measured after the gel extraction step described above) topped up with up with dH_2O to a final volume of 50 µl. The reaction was incubated for 14 hours at room temperature and for 30 minutes at 4^{o}C. Then beads were washed twice with 200 µl 1x Tween buffer, four times with 200 µl of 1x no Tween buffer, with 200 µl of 1x NEB2 buffer and re-suspended in 50 µl of 1x NEB2 buffer. The Hi-C library was amplified by the PCR reaction with the Heculase Fusion II fusion DNA polymerase (Agilent) protocol for DNA targets 1-10 kb and was set up as follows: 10 µl of the above beads dilution, 10 µl of 5x Heculase Fusion II buffer, 2.5 µl DMSO, 1 mM dNTP mix, 25 µM Illumina PE primer 1.0 and 2.0, 1 µl of Heculase Fusion II Polymerase topped up with dH_2O to a final volume of 50 µl. The reaction conditions were as follows: 98^{o}C for 2 minutes (1 cycle); 98^{o}C for 20 seconds, 65^{o}C for 30 seconds, 72^{o}C for 30 seconds (8 cycles); 72^{o}C for 5 minutes (1 cycle). Then, the PCR product was purified into 20 µl dH_2O using AMPureXP beads according to the manufacturer's instructions. Finally, 1 µl of the sample was evaluated on a Bioanalyzer chip using the Agilent DNA 1000 kit. In optimal conditions, the Bioanalyzer profile of each library should be between 300 to 500 bp sizes (**Figure 36**).

Based on the above Bioanalyzer results, each library was diluted to 10 nM and the libraries were sequenced in the HiSeq 2500 Ultra-High-Throughput Sequencing System by UCL Genomics based at the UCL Cancer Institute. Comparable number of paired-end 50 bp reads between datasets were generated to be aligned by Dr C. Barrington to the mouse genome (mm10) and inputted to the Hi-C analysis pipeline to generate the Hi-C contact maps (Yaffe and Tanay, 2011) (library information in **Table 4** in the APPENDIX).

The Hi-C contact maps allowed for the visualisation of a certain genomic region of variable sizes. The diagonal in the map depicts the linear space of the chromosome or of a certain region of interest on a particular chromosome and each each pixel-square in the map represents the normalised value (observed/expected) of frequency of contacts between to restriction fragments in colour-code at a certain distance *in cis*. This linear distance on the chromosome of the two fragments can be defined if two lines are drawn from this pixel to the diagonal, where these lines form two sides of an equilateral triangle with the linear distance of the fragments.

APPENDIX

Gene Symbol	Forward Primer 5'-3'	Reverse Primer 5'-3'
Tbp	CTGAAGAAAGGGAGAATCATGG	TGTCTTTGTTGCTCTTCCAAAA
Ubc	AGGAGGCTGATGAAGGAGCTTGA	TGGTTTGAATGGATACTCTGCTGGA
Gfap	TGCAAGAGACAGAGGAGTGGTA	CGATAGTCGTTAGCTTCGTGCT
Aqp4	CAGCTGTGATTCCAAACGAA	GCCCAGTTTCCCATGATAAC
Egfr	ATGTGCAAAGGAATTACGACCT	CTAAGCCCAGTTCTGTTTGTCC
Blbp	AAGGATGGTAGATGCTTTCTGC	TGCCACCTTCCTGACTGATAAT
Ccnd1	TGAAGGAGACCATTCCCTTG	CTGGCATTTTGGAGAGGAAG
Cdkn2d	GTTCTTGGTCACTGTGAGGATTCAG	CCATCATCATCACCTGGTCCAG
Rad21	ACGAAGAAGCTTTTGCGTTG	CGCTAAGCTGGGCTCTAATG
Oct4	CCCAAGGTGATCCTCTTCTGCTT	GAGAAGGTGGAACCAACTCCCG
Nanog	GCATCTTCTGCTTCCTGGCAA	GAACTATTCTTGCTTACAAGGGTCTG
Sox2	TGGACTGCGAACTGGAGAAGG	CGCCCGGAGTCTAGCTCTAAATATT
Rex1	GAGCTGAACTCCTAGCCGCCTAGATT	TTTGGTCAGTGGTATTTGGGGACA
Glast	GTCGCGGTGATAATGTGGTA	AATCTTCCCTGCGATCAAGA
Olig2	CCAGAGCCAGGTTCTCCTC	CCCCAGGGATGATCTAAGC
Mash1	TCTTAGCCCAGAGGAACAAGAG	GAATGCAGAGACACTGTTGGAG
Fgfr3	CCTGGCGGGCAATTCTAT	GCCACCACCAGGATGAAG
Slc6a8	AGCCACTGGTCTACAACAACA	CTCTGCCATGGTTCCTTTTG
Slc1a2	GGAAGAAGAACGACGAGGTG	GCCTCGTTCATGGTTTCATT
Ctcf	GTGGCCAAATTTCATTGTCC	CGCTCATGAAACACAGCATC
Pcdha1	CTGGTCATACTCGCAGCAGA	TGCTCTTAGCGAGGCAGAGT
Pcdha2	CTGGTCATACTCGCAGCAGA	TGCTCTTAGCGAGGCAGAGT

Pcdha3	TGCTGTCGAGCTATGTGTCC	GTCGTTCTCATCCAGCACAA
Pcdha4	CTACGCGCATACTGGATGAA	CGTTGACATCCACCAGTGAC
Pcdha5	AAAATACACCCAACGGGACA	CGTAATCCAGTTCGCCCTTA
Pcdha6	CCAGCGAATACTTTGGGCTA	CAGGAGCATCCTCTCTGTCC
Pcdha8	TCACCGTGCTGGTGTCTCTA	CAGATGGCGATGATCAGGTA
Pcdha9	GGTGGAGGTGAGGGACATTA	AGGCGCCTTCTAGTGGAAAT
Pcdha10	AGTGGGTTCAGGGCATGTAG	CAGAGCGCGAGTGATACTGA
Pcdha11	GGGAGCTGGTCCTACTCACA	TGCTCTTAGCGAGGCAGAGT
Pcdha12	TCCTTCAGACGGCCTGTATC	GCCTCCACGGGTATCTCATA
Pcdhac1	GCCTAAGCATTAGCCAGCAC	GTTGGTGTCCACCACAGTGA
Pcdhac2	ATGTGAATGACAACGCTCCA	GGAGTTTCCAAAGCCATTCA
Pcdhcon1-2	CGCTACTCTGCCTCGCTAAG	GGATACTGTTGGCCACTGCT

Table 1. Sequences of the primers used in the real time PCR assays. The genes and the primer sequences (5'-3') used to investigate the mRNA expression. The stock concentration of each primer was 10μM (Sigma) and they were all used for the Roche SYBR green qPCR assays.

Genomic location	Forward Primer	Forward Primer
Chr 9:44574238	GCGGCTACTGTGTTCTAGCC	CTCCAGCGTCTTTTTGGAAG
XIST 7N	TGGCTTGTACTTCCAGATCAT	AATGTAAAGCAAGCTAGTACGCA
Chr 1:22250000	TGGTAAAGACAGTTTCCGAGGT	AACAAATGTAGTGGGGCATTTC
Chr 16:23980000	TTGAGATCCACAGAGCTCACAT	AGGAAGGGTCCTTGGTTCTTAG
Chr 10:2448	CCACATAACTCCAGTTCATCCA	CTTCCTTCTTGCTCTTTTGCAT

Table 2. Sequences of the primers used in the real time PCR assays for ChIP-qPCR. The chromosome locations and the primer sequences (5'-3') used to investigate Rad21 binding enrichment. The stock concentration of each primer was 10μM (Sigma) and they were all used for the Roche SYBR green ChIP-qPCR assays.

Genomic location of Rad21 peak	Primers to Rad21 peak (bp)	Reading primer	Non-reading primer
Chr 15: 54928695-985	123	GATGGAAACTGGGAATGATC	CCTGGAACTCTCTGTGTTTT
Chr 15: 54944702-980	843	GGCTACCCTGGCTAGTGATC	ATTTGAATTTGAGGGTATGC
Chr 18: 37180253-494	1,523	TGTAAAGCAGCAGCGTGATC	TGTGTCACAAACTGGAAAAA
Chr 18: 37377915-8314	432	GTATTGTAAAGTGGTCGATC	AAGTTGTAGTGGCAGGATTG
Chr 18: 38039914-40097	432	GATTCTGAACCCGACAGATC	TTTATCAGTGTTGGAGTAGCTT

Table 3. Sequences of the primers used in the 4C-Seq. The chromosome locations of the Rad21 ChIP-Seq peak that 4C-Seq baits were designed to, their distance from the primers and the primer sequences (5'-3') used to investigate the frequency of chromatin contacts from the Rad21 peaks as 4C-Seq baits. The stock concentration of each primer was 10 μM (Sigma) and they were all used for the Roche Expand Long Template PCR at the library preparation step.

Experiment	Cell type	Library	Total reads	Aligned reads
RNA-Seq	NS	*Rad21*$^{Lox/Lox}$ Rep1	197,575,770	171,847,072
		Rad21$^{\Delta/\Delta}$ Rep1	196,824,172	172,370,348
		Rad21$^{\Delta/\Delta}$ Rep2	196,389,640	173,230,450
	AST	*Rad21*$^{Lox/Lox}$ 72h Rep1	64,617,898	44,991,352
		Rad21$^{Lox/Lox}$ 72h Rep2	128,034,010	99,372,356
		Rad21$^{\Delta/\Delta}$ 72h Rep1	171,500,112	161,007,519
		Rad21$^{\Delta/\Delta}$ 72h Rep2	161,582,478	151,628,664
		Rad21$^{Lox/Lox}$ 96h Rep1	96,652,864	73,299,209
		Rad21$^{Lox/Lox}$ 96h Rep2	157,608,996	142,134,841
		Rad21$^{\Delta/\Delta}$ 96h Rep1	135,376,776	108,848,912
		Rad21$^{\Delta/\Delta}$ 96h Rep2	162,328,468	140,294,262
ChIP-Seq	AST	*Rad21*$^{Lox/Lox}$ Rep1	17,965,132	15,791,351

(Rad21)		*Rad21*$^{Lox/Lox}$ Rep2	16,148,475	14,057,247
		Rad21$^{\Delta/\Delta}$ Rep1	17,238,004	13,928,307
		Rad21$^{\Delta/\Delta}$ Rep2	14,650,933	11,544,935
		Rad21$^{\Delta/\Delta}$+siCTCF Rep1	15,871,143	12,811,186
		Rad21$^{\Delta/\Delta}$+siCTCF Rep2	13,720,983	10,935,623
		Rad21$^{\Delta/\Delta}$+siStag1 Rep1	15,797,332	11,587,343
		Rad21$^{\Delta/\Delta}$+siStag1 Rep2	12,526,791	8,580,851
Hi-C	AST	*Rad21*$^{Lox/Lox}$	1,486,886,140	1,089,408,144
		Rad21$^{\Delta/\Delta}$	1,481,910,324	1,075,343,153
		Rad21$^{\Delta/\Delta}$+siCTCF	726,282,526	527,498,617
		Rad21$^{\Delta/\Delta}$+siStag1	790,839,396	572,000,000
		Rad21$^{\Delta/\Delta}$+siStag2	1,019,219,572	735,000,000

Table 4. Read statistics from the RNA-Seq, ChIP-Seq, Hi-C datasets. The RNA-Seq libraries (strand-specific, paired-end) were sequenced on the Genome Analyzer II in Illumina Inc (San Diego), the ChIP-Seq libraries (single-end, 75 bp) were sequenced at the MiSeq and the Hi-C libraries (paired-end, 55 bp) at the HiSeq 2500 platform at the UCL Cancer Institute. Hi-C dataset for *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ ASTs are from Sofueva *et al.* 2013.

GENE	*Rad21*$^{Lox/Lox}$ 72h	*Rad21*$^{\Delta/\Delta}$ 72h	z-score	*Rad21*$^{Lox/Lox}$ 96h	*Rad21*$^{\Delta/\Delta}$ 96h	z-score
Gm13086	8.81148	98.5163	5.401992241	11.8397	163.853	6.700495939
Gm12896	2.36206	34.8014	4.393411014	82.1765	141.015	1.645215459
Mest	2.83287	18.2394	3.398818291	3.01778	24.084	3.889358405
Trnp1	1.79835	17.117	3.733834528	2.09028	21.7046	4.058876085
Defb1	1.61934	10.6937	3.327630071	2.85344	29.1799	4.326688744
Rnf128	29.4029	399.714	7.94524525	20.065	417.774	7.749511085
Mt1	12.8972	1467.12	14.47680019	22.4887	1483.58	10.61899005
Iapp	0.119478	2.06734	3.182294222	0.378566	3.00049	2.98536858

Speer4e	0.13384	2.0213	3.029364688	0.226124	4.42014	3.892522929
Npr3	0.0968771	3.37565	3.969246799	0.103743	4.2716	3.731825378
BC061195	0.0302345	2.97502	3.820358403	0.137195	6.81096	3.925473441
Samd12	0.238474	1.23587	2.483642587	0.413658	2.03968	2.366705748
Ccdc114	0.24689	1.14786	2.337618029	0.46448	3.04139	2.736422308
Dlk1	0.151372	0.844544	2.273679475	0.136416	2.34986	2.829700107
U3	0.121821	0.617459	1.801978385	0.495544	2.41495	2.351414937
Palm3	0.336436	4.00718	3.758285507	0.741746	7.07671	3.538341176
Myoc	0.257904	3.55419	3.799207284	0.249788	3.88767	3.604098176
1700003M07Rik	0.408371	1.3001	1.862910452	0.501229	1.46139	1.681831603
Myh4,Myh8	0.933143	7.46827	3.358213052	1.27053	9.99609	3.42941743
Igf1	0.580717	10.0001	4.145074262	0.829627	11.232	4.039837328
Col8a1	0.485236	15.3176	4.835037319	0.807769	19.5105	4.869319406
Slc26a9	0.0608132	0.471801	1.968184692	0.081763	0.864865	1.830484574
AK129341	0.0506297	0.430783	2.062083242	0.0575654	0.741352	2.00250157
Lrrc3	0.249716	1.1144	2.282515599	0.231226	1.15364	2.164228653
Lhx6	0.0569908	0.308604	1.602437952	0.0195456	0.510736	2.321644928
Tmem119	0.418853	2.62746	2.741595413	0.180128	1.36091	2.351753937
Vmn1r42	0.0546975	0.378378	1.851543758	0.0671455	0.541621	1.593579416
Col6a3	0.0551425	0.483308	2.092259182	0.0656358	0.499253	1.542261606
Kcnj15	0.103081	0.645534	2.039748887	0.0692834	0.902711	2.012678329
Speer4d	0.102199	0.561817	1.893256694	0.0545315	1.37136	2.58743629
Ada	0.131018	0.867638	2.10256561	0.0771176	1.16663	2.143213619
Acer2	0.22664	1.04526	2.326784881	0.225565	1.51636	2.541611178
Spna1	0.0577946	0.828453	2.592625199	0.0703837	1.1145	2.183119493
Grip1	0.0450979	0.601088	2.518618577	0.437009	1.6398	2.010688361

Cdh26	0.041571	0.710329	2.771297747	0.161693	2.7059	3.269106964
Cyp2f2	0.0222837	0.667915	2.719884653	0.0451997	2.0694	3.111116037
Igf2bp1	0.0140265	0.734597	3.236525234	0.00264511	4.11413	5.661322962
Gprc5a	0.0123218	0.135148	1.789207414	0.0200721	0.213128	1.585657432
Pkp1	0.0108971	0.136081	1.909328495	0.0135211	0.263226	2.081087893
Ptprt	0.00821497	0.154839	2.290442455	0.00844722	0.294947	2.558536154
C230038L03Rik	0.00991746	0.111136	1.809071959	0.00857343	0.409006	2.813610066
Ndst3	0.0171898	0.166387	1.673486469	0.0116311	0.173445	1.863216336
Col3a1	0.0208555	0.20002	1.664968907	0.0181734	0.89172	2.836600984
Ces2e	0.301736	1.03501	2.043618196	0.141412	5.01215	4.134784832
Scn4b	0.0052858	0.107401	2.359983912	0.0167228	0.246456	1.85359901
Gpr64	0.00506897	0.0461808	1.617409369	0.0046287	0.0857873	2.040905334
Gm12198	0.0923883	0.0162902	-1.973627034	0.0481436	0.00746377	-1.860506835
Prdx6b	0.0964527	0.0193958	-1.825862309	0.0217277	0.00498973	-1.547578148
5830403L16Rik	0.108221	0.0124842	-2.450594849	0.0348876	0.00492918	-1.94457073
0610040F04Rik	0.110221	0.0210584	-1.883411797	0.0748873	0.00779949	-2.20820126
Serpind1	0.139388	0.0210789	-2.164460328	0.114124	0.0079789	-2.8711939
Fgd2	0.0969444	0.0118638	-2.384193459	0.117982	0.0030143	-3.913338521
Icosl	3.33109	1.15271	-1.533753246	4.20406	1.18189	-2.300470937
Grin2c	0.204522	0.00977981	-3.819104734	0.0864308	0.00699194	-2.720154913
Cacna2d4	0.180361	0.0241657	-2.313344338	0.152318	0.0134702	-2.778889549
Gm15641	0.195101	0.0260372	-2.318221511	0.0782451	0.0116786	-1.893686121
1700066J2Rik	0.147765	0.024192	-2.06680809	0.0729887	0.010902	-1.893048308
Kcnj9	0.167166	0.0407379	-1.577520756	0.11581	0.0083676	-2.837133587
Slc25a41	0.0484402	0.00640911	-2.172982806	0.0578875	0.00601631	-2.210037971

Afp	0.0634709	0.00688368	-2.375210253	0.0507024	0.00640325	-2.039721399
Gm5648	0.0682871	0.00928488	-2.145239106	0.0285772	0.00621498	-1.592070941
Rgsl1	0.0572239	0.00403701	-2.812622879	0.0575989	0.0017844	-3.267978878
BC024139	0.325101	0.0627286	-1.920562783	0.280647	0.0166723	-3.361822593
Gm12009	0.625937	0.055816	-2.901388677	0.124484	0.0142289	-2.362292557
Gm2636	0.341801	0.0270234	-3.149525141	0.153884	0.0176729	-2.47744154
Vwa3b	0.377673	0.0642599	-2.093809727	0.338788	0.0329508	-2.723689747
Gm7932	0.541635	0.0633163	-2.53892304	1.45356	0.0330654	-5.522580682
Vtcn1	0.340156	0.0439137	-2.474100495	0.60068	0.0396247	-3.379559478
Gm14556	1.06824	0.242929	-1.837890289	0.266694	0.0642012	-1.672351582
Wnt5b	1.82706	0.156845	-3.373984658	1.5022	0.0519849	-4.880125862
Slpi	1.5912	0.285665	-2.388067978	0.590734	0.0253367	-3.964983792
Gm12891	0.916581	0.0778744	-3.274579662	0.806232	0.0584968	-3.45268526
Gm12460	1.03968	0.100033	-3.093073563	0.19175	0.0342852	-1.966782514
Atp5o	0.594341	0.140189	-1.599103199	0.482797	0.0273995	-3.452689135
Fabp4	14.5039	4.36269	-3.731948452	16.8257	0.529234	-8.411807692
Lpl	19.8603	3.99584	-4.618271887	16.8136	1.25936	-6.252930428
Kcnj8	11.6163	1.66264	-5.106401597	12.8065	0.737547	-6.968182758
Opn4	15.0501	5.45982	-3.157159202	14.532	1.72544	-5.164171867
Gm14287	90.5907	5.15413	-5.511812585	6.84835	1.89348	-2.552419392
Gm11572	58.384	5.69616	-6.719529671	1.98497	0.551881	-1.889321597
Mmp11	165.786	37.9115	-2.778655466	47.812	14.1363	-3.063836724
Smpd3	4.07774	0.873131	-2.343670822	3.21328	0.238736	-4.360072087
Gm6166	4.81436	0.693743	-3.723515742	4.54199	0.405709	-4.552422407
Adora1	19.5058	8.46308	-2.389340906	20.8429	8.45271	-2.050222962
Gpr37l1	3.481	0.459668	-3.159205351	4.41941	0.288557	-5.16794511

Nrtn	1.77017	0.26018	-2.562199257	0.955361	0.194693	-1.972887044
Lrtm2	1.49727	0.227043	-2.655750619	0.877486	0.158592	-2.144921419
Gm16379	9.55139	3.18239	-2.859613196	0.416325	0.0351264	-2.93579187
Gm11512	20.6419	0.839963	-9.253964935	0.308336	0.0290344	-2.765563711
Gm14472	3.90165	0.881563	-2.252975197	0.840607	0.0518264	-3.685996291
Mxra8	5.65332	1.01894	-3.273051058	0.296025	0.0450997	-2.142769069

Table 5. The expression of de-regulated genes in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ ASTs at 72 and 96 hours post Cre-expressing adenovirus-infections. The RNA-Seq expression data (FPKM values) and z-scores of the top up- and down-regulated genes in *Rad21*$^{Lox/Lox}$ and *Rad21*$^{\Delta/\Delta}$ ASTs at 72 and 96 hours post-infections used for the heatmap in Figure 21.

REFERENCES

Anderson, D.E., Losada, A., Erickson, H.P., and Hirano, T. (2002). Condensin and cohesin display different arm conformations with characteristic hinge angles. The Journal of Cell Biology *156*, 419-424.

Apostolou, E., Ferrari, F., Walsh, R.M., Bar-Nur, O., Stadtfeld, M., Cheloufi, S., Stuart, H.T., Polo, J.M., Ohsumi, T.K., Borowsky, M.L.*, et al.* (2013). Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. Cell stem cell *12*, 699-712.

Arumugam, P., Gruber, S., Tanaka, K., Haering, C.H., Mechtler, K., and Nasmyth, K. (2003). ATP Hydrolysis Is Required for Cohesin's Association with Chromosomes. Current Biology *13*, 1941-1953.

Arumugam, P., Nishino, T., Haering, C.H., Gruber, S., and Nasmyth, K. (2006). Cohesin's ATPase Activity Is Stimulated by the C-Terminal Winged-Helix Domain of Its Kleisin Subunit. Current Biology *16*, 1998-2008.

Balbas-Martinez, C., Sagrera, A., Carrillo-de-Santa-Pau, E., Earl, J., Marquez, M., Vazquez, M., Lapi, E., Castro-Giner, F., Beltran, S., Bayes, M.*, et al.* (2013). Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. Nat Genet *45*, 1464-1469.

Beckouet, F., Hu, B., Roig, M.B., Sutani, T., Komata, M., Uluocak, P., Katis, V.L., Shirahige, K., and Nasmyth, K. (2010). An Smc3 acetylation cycle is essential for establishment of sister chromatid cohesion. Mol Cell *39*, 689-699.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482-485.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387-396.

Ben-Shahar, T.R., Heeger, S., Lehane, C., East, P., Flynn, H., Skehel, M., and Uhlmann, F. (2008). Eco1-Dependent Cohesin Acetylation During Establishment of Sister Chromatid Cohesion. Science *321*, 563-566.

Bickmore, W. (1999). Fluorescence in situ hybridization analysis of chromosome and chromatin structure. Methods in enzymology *304*, 650-662.

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Muller, S., Eils, R., Cremer, C., Speicher, M.R.*, et al.* (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS biology *3*, e157.

Brown, J.M., Leach, J., Reittie, J.E., Atzberger, A., Lee-Prudhoe, J., Wood, W.G., Higgs, D.R., Iborra, F.J., and Buckle, V.J. (2006). Coregulated human globin genes are frequently in spatial proximity when active. J Cell Biol *172*, 177-187.

Buheitel, J., and Stemmann, O. (2013). Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3-Scc1 gate. The EMBO journal *32*, 666-676.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327-339.

Canudas, S., and Smith, S. (2009). Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. J Cell Biol *187*, 165-173.

Carretero, M., Ruiz-Torres, M., Rodriguez-Corsino, M., Barthelemy, I., and Losada, A. (2013). Pds5B is required for cohesion establishment and Aurora B accumulation at centromeres. The EMBO journal *32*, 2938-2949.

Chan, K.L., Gligoris, T., Upcher, W., Kato, Y., Shirahige, K., Nasmyth, K., and Beckouet, F. (2013). Pds5 promotes and protects cohesin acetylation. Proc Natl Acad Sci U S A *110*, 13020-13025.

Chan, K.L., Roig, M.B., Hu, B., Beckouet, F., Metson, J., and Nasmyth, K. (2012). Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. Cell *150*, 961-974.

Chien, R., Zeng, W., Kawauchi, S., Bender, M.A., Santos, R., Gregson, H.C., Schmiesing, J.A., Newkirk, D.A., Kong, X., Ball, A.R., Jr.*, et al.* (2011). Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. The Journal of biological chemistry *286*, 17870-17878.

Chung, J.H., Whiteley, M., and Felsenfeld, G. (1993). A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. Cell *74*, 505-514.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol Cell *5*, 243-254.

Conti, L., and Cattaneo, E. (2010). Neural stem cell systems: physiological players or in vitro entities? Nat Rev Neurosci *11*, 176-187.

Conti, L., Pollard, S.M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.L., Cattaneo, E.*, et al.* (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. PLoS biology *3*, e283.

Cook, A.J., Gurard-Levin, Z.A., Vassias, I., and Almouzni, G. (2011). A specific function for the histone chaperone NASP to fine-tune a reservoir of soluble H3-H4 in the histone supply chain. Mol Cell *44*, 918-927.

Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J., and Meyer, B.J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature *523*, 240-244.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nature reviews Genetics *2*, 292-301.

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., and Bickmore, W.A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. J Cell Biol *145*, 1119-1131.

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res *19*, 24-32.

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes Dev *26*, 11-24.

Deardorff, M.A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pie, J., Gil-Rodriguez, C., Arnedo, M., Loeys, B., Kline, A.D.*, et al.* (2007). Mutations in cohesin complex

members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. American journal of human genetics *80*, 485-494.

Deardorff, M.A., Wilde, J.J., Albrecht, M., Dickinson, E., Tennstedt, S., Braunholz, D., Monnich, M., Yan, Y., Xu, W., Gil-Rodriguez, M.C.*, et al.* (2012). RAD21 mutations cause a human cohesinopathy. American journal of human genetics *90*, 1014-1027.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

DeMare, L.E., Leng, J., Cotney, J., Reilly, S.K., Yin, J., Sarro, R., and Noonan, J.P. (2013). The genomic landscape of cohesin-associated chromatin interactions. Genome Res *23*, 1224-1234.

Dimova, D.K., and Dyson, N.J. (2005). The E2F transcriptional network: old acquaintances with new faces. Oncogene *24*, 2810-2826.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Dorsett, D., Eissenberg, J.C., Misulovin, Z., Martens, A., Redding, B., and McKim, K. (2005). Effects of sister chromatid cohesion proteins on cut gene expression during wing development in Drosophila. Development *132*, 4743-4753.

Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C.*, et al.* (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res *16*, 1299-1309.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. Nature *465*, 363-367.

Eichinger, C.S., Kurze, A., Oliveira, R.A., and Nasmyth, K. (2013). Disengaging the Smc3/kleisin interface releases cohesin from Drosophila chromosomes during interphase and mitosis. The EMBO journal *32*, 656-665.

El-Kady, A., and Klenova, E. (2005). Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. FEBS Lett *579*, 1424-1434.

Farrell, C.M., West, A.G., and Felsenfeld, G. (2002). Conserved CTCF insulator elements flank the mouse and human beta-globin loci. Mol Cell Biol *22*, 3820-3831.

Faure, A.J., Schmidt, D., Watt, S., Schwalie, P.C., Wilson, M.D., Xu, H., Ramsay, R.G., Odom, D.T., and Flicek, P. (2012). Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. Genome Res *22*, 2163-2175.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V.V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol *16*, 2802-2813.

Fraser, P., and Bickmore, W. (2007). Nuclear organization of the genome and the potential for gene regulation. Nature *447*, 413-417.

Fraser, J., Williamson, I., Bickmore, W., Dostie, J. (2015). An Overview of Genome Organization and How We Got There: from FISH to Hi-C. Microbiol. Mol. Biol. Rev. *79*, 347-372.

Fraser, P., Pruzina, S., Antoniou, M., and Grosveld, F. (1993). Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. Genes Dev *7*, 106-113.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H.*, et al.* (2009a). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58-64.

Fullwood, M.J., Wei, C.L., Liu, E.T., and Ruan, Y. (2009b). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. Genome Res *19*, 521-532.

Furey, T.S. (2012). ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nature reviews Genetics *13*, 840-852.

Gandhi, R., Gillespie, P.J., and Hirano, T. (2006). Human Wapl Is a Cohesin-Binding Protein that Promotes Sister-Chromatid Resolution in Mitotic Prophase. Current Biology *16*, 2406-2417.

Gard, S., Light, W., Xiong, B., Bose, T., McNairn, A.J., Harris, B., Fleharty, B., Seidel, C., Brickner, J.H., and Gerton, J.L. (2009). Cohesinopathy mutations disrupt the subnuclear organization of chromatin. The Journal of Cell Biology *187*, 455-462.

Gause, M., Misulovin, Z., Bilyeu, A., and Dorsett, D. (2010). Dosage-sensitive regulation of cohesin chromosome binding and dynamics by Nipped-B, Pds5, and Wapl. Mol Cell Biol *30*, 4940-4951.

Gerlich, D., Koch, B., Dupeux, F., Peters, J.M., and Ellenberg, J. (2006). Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. Current biology : CB *16*, 1571-1578.

Gillespie, P.J., and Hirano, T. (2004). Scc2 couples replication licensing to sister chromatid cohesion in Xenopus egg extracts. Current biology : CB *14*, 1598-1603.

Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell *157*, 950-963.

Gligoris, T.G., Scheinost, J.C., Bürmann, F., Petela, N., Chan, K.-L., Uluocak, P., Beckouët, F., Gruber, S., Nasmyth, K., and Löwe, J. (2014). Closing the cohesin ring: Structure and function of its Smc3-kleisin interface. Science *346*, 963-967.

Goldman, R.D., Shumaker, D.K., Erdos, M.R., Eriksson, M., Goldman, A.E., Gordon, L.B., Gruenbaum, Y., Khuon, S., Mendez, M., Varga, R.*, et al.* (2004). Accumulation of mutant lamin A causes progressive changes in nuclear architecture in Hutchinson-Gilford progeria syndrome. Proc Natl Acad Sci U S A *101*, 8963-8968.

Griffith, J., Hochschild, A., and Ptashne, M. (1986). DNA loops induced by cooperative binding of lambda repressor. Nature *322*, 750-752.

Grosveld, F., Antoniou, M., van Assendelft, G.B., de Boer, E., Hurst, J., Kollias, G., MacFarlane, F., and Wrighton, N. (1987). The regulation of expression of human beta-globin genes. Progress in clinical and biological research *251*, 133-144.

Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K., and Nasmyth, K. (2006). Evidence that Loading of Cohesin Onto Chromosomes Involves Opening of Its SMC Hinge. Cell *127*, 523-537.

Gruber, S., Haering, C.H., and Nasmyth, K. (2003). Chromosomal cohesin forms a ring. Cell *112*, 765-777.

Guacci, V., Koshland, D., and Strunnikov, A. (1997). A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. Cell *91*, 47-57.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W*., et al.* (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948-951.

Guillou, E., Ibarra, A., Coulon, V., Casado-Vela, J., Rico, D., Casal, I., Schwob, E., Losada, A., and Mendez, J. (2010). Cohesin organizes chromatin loops at DNA replication factories. Genes & development *24*, 2812-2822.

Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K.E., Li, W., Myers, R.M., Maniatis, T., and Wu, Q. (2012). CTCF/cohesin-mediated DNA looping is required for protocadherin alpha promoter choice. Proc Natl Acad Sci U S A *109*, 21081-21086.

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature *460*, 410-413.

Haering, C.H., Farcas, A.M., Arumugam, P., Metson, J., and Nasmyth, K. (2008). The cohesin ring concatenates sister DNA molecules. Nature *454*, 297-301.

Haering, C.H., Lowe, J., Hochwagen, A., and Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. Mol Cell *9*, 773-788.

Haering, C.H., Schoffnegger, D., Nishino, T., Helmhart, W., Nasmyth, K., and Lowe, J. (2004). Structure and stability of cohesin's Smc1-kleisin interaction. Mol Cell *15*, 951-964.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W., Ye, C., Ping, J.L., Mulawadi, F*., et al.* (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet *43*, 630-638.

Hauf, S., Roitinger, E., Koch, B., Dittrich, C.M., Mechtler, K., and Peters, J.M. (2005). Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. PLoS biology *3*, e69.

Hauf, S., Waizenegger, I.C., and Peters, J.M. (2001). Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. Science *293*, 1320-1323.

Heidinger-Pauli, J.M., Mert, O., Davenport, C., Guacci, V., and Koshland, D. (2010). Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair. Current biology : CB *20*, 957-963.

Herold, M., Bartkuhn, M., and Renkawitz, R. (2012). CTCF: insights into insulator function during development. Development *139*, 1045-1057.

Hirano, M., Anderson, D.E., Erickson, H.P., and Hirano, T. (2001). Bimodal activation of SMC ATPase by intra- and inter-molecular interactions. The EMBO journal *20*, 3238-3250.

Hirano, M., and Hirano, T. (2002). Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. The EMBO journal *21*, 5733-5744.

Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. Nature reviews Molecular cell biology *7*, 311-322.

Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N., and Yagi, T. (2012). CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. Cell reports *2*, 345-357.

Horsfield, J.A., Anagnostou, S.H., Hu, J.K., Cho, K.H., Geisler, R., Lieschke, G., Crosier, K.E., and Crosier, P.S. (2007). Cohesin-dependent regulation of Runx genes. Development *134*, 2639-2649.

Hou, F., and Zou, H. (2005). Two human orthologues of Eco1/Ctf7 acetyltransferases are both required for proper sister-chromatid cohesion. Molecular biology of the cell *16*, 3908-3918.

Hu, B., Itoh, T., Mishra, A., Katoh, Y., Chan, K.-L., Upcher, W., Godlee, C., Roig, M.B., Shirahige, K., and Nasmyth, K. (2011). ATP Hydrolysis Is Required for Relocating Cohesin from Sites Occupied by Its Scc2/4 Loading Complex. Current Biology *21*, 12-24.

Huis in 't Veld, P.J., Herzog, F., Ladurner, R., Davidson, I.F., Piric, S., Kreidl, E., Bhaskara, V., Aebersold, R., and Peters, J.-M. (2014). Characterization of a DNA exit gate in the human cohesin ring. Science *346*, 968-972.

Ivanov, D., and Nasmyth, K. (2005). A topological interaction between cohesin rings and a circular minichromosome. Cell *122*, 849-860.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G.*, et al.* (2011). Spatio-temporal transcriptome of the human brain. Nature *478*, 483-489.

Kawauchi, S., Calof, A.L., Santos, R., Lopez-Burks, M.E., Young, C.M., Hoang, M.P., Chua, A., Lao, T., Lechner, M.S., Daniel, J.A.*, et al.* (2009). Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. PLoS Genet *5*, e1000650.

Kehayova, P., Monahan, K., Chen, W., and Maniatis, T. (2011). Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. Proc Natl Acad Sci U S A *108*, 17195-17200.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231-1245.

Kitajima, T.S., Sakuno, T., Ishiguro, K.-i., Iemura, S.-i., Natsume, T., Kawashima, S.A., and Watanabe, Y. (2006). Shugoshin collaborates with protein phosphatase 2A to protect cohesin. Nature *441*, 46-52.

Klenova, E.M., Chernukhin, I.V., El-Kady, A., Lee, R.E., Pugacheva, E.M., Loukinov, D.I., Goodwin, G.H., Delgado, D., Filippova, G.N., Leon, J.*, et al.* (2001). Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. Mol Cell Biol *21*, 2221-2234.

Kon, A., Shih, L.Y., Minamino, M., Sanada, M., Shiraishi, Y., Nagata, Y., Yoshida, K., Okuno, Y., Bando, M., Nakato, R.*, et al.* (2013). Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nat Genet *45*, 1232-1237.

Kong, X., Ball, A.R., Jr., Pham, H.X., Zeng, W., Chen, H.Y., Schmiesing, J.A., Kim, J.S., Berns, M., and Yokomori, K. (2014). Distinct functions of human cohesin-SA1 and cohesin-SA2 in double-strand break repair. Mol Cell Biol *34*, 685-698.

Koshland, D., and Hartwell, L.H. (1987). The structure of sister minichromosome DNA before anaphase in Saccharomyces cerevisiae. Science *238*, 1713-1716.

Krantz, I.D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L.A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C.A.*, et al.* (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nat Genet *36*, 631-635.

Kueng, S., Hegemann, B., Peters, B.H., Lipp, J.J., Schleiffer, A., Mechtler, K., and Peters, J.M. (2006). Wapl controls the dynamic association of cohesin with chromatin. Cell *127*, 955-967.

Kulemzina, I., Schumacher, M.R., Verma, V., Reiter, J., Metzler, J., Failla, A.V., Lanz, C., Sreedharan, V.T., Ratsch, G., and Ivanov, D. (2012). Cohesin rings devoid of Scc3 and Pds5 maintain their stable association with the DNA. PLoS Genet *8*, e1002856.

Kumada, K., Yao, R., Kawaguchi, T., Karasawa, M., Hoshikawa, Y., Ichikawa, K., Sugitani, Y., Imoto, I., Inazawa, J., Sugawara, M.*, et al.* (2006). The selective continued linkage of centromeres from mitosis to interphase in the absence of mammalian separase. J Cell Biol *172*, 835-846.

Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc Natl Acad Sci U S A *103*, 10684-10689.

Kurze, A., Michie, K.A., Dixon, S.E., Mishra, A., Itoh, T., Khalid, S., Strmecki, L., Shirahige, K., Haering, C.H., Lowe, J.*, et al.* (2011). A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion. The EMBO journal *30*, 364-378.

Ladurner, R., Bhaskara, V., Huis in 't Veld, P.J., Davidson, I.F., Kreidl, E., Petzold, G., and Peters, J.M. (2014). Cohesin's ATPase activity couples cohesin loading onto DNA with Smc3 acetylation. Current biology : CB *24*, 2228-2237.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S.*, et al.* (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*, 943-949.

Laugsch, M., Seebach, J., Schnittler, H., and Jessberger, R. (2013). Imbalance of SMC1 and SMC3 cohesins causes specific and distinct effects. PLoS One *8*, e65149.

Leiserson, M.D., Vandin, F., Wu, H.T., Dobson, J.R., Eldridge, J.V., Thomas, J.L., Papoutsaki, A., Kim, Y., Niu, B., McLellan, M.*, et al.* (2015). Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nat Genet *47*, 106-114.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature *430*, 573-578.

Lengronne, A., McIntyre, J., Katou, Y., Kanoh, Y., Hopfner, K.-P., Shirahige, K., and Uhlmann, F. (2006). Establishment of Sister Chromatid Cohesion at the S. cerevisiae Replication Fork. Molecular Cell *23*, 787-799.

Lettice, L.A., Horikoshi, T., Heaney, S.J., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N.*, et al.* (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc Natl Acad Sci U S A *99*, 7548-7553.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Liu, H., Rankin, S., and Yu, H. (2013). Phosphorylation-enabled binding of SGO1-PP2A to cohesin protects sororin and centromeric cohesion during mitosis. Nat Cell Biol *15*, 40-49.

Liu, J., and Krantz, I.D. (2009). Cornelia de Lange syndrome, cohesin, and beyond. Clin Genet *76*, 303-314.

Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M.A., Clark, D., Kaur, M., Tandy, S., Kondoh, T., Rappaport, E.*, et al.* (2009). Transcriptional dysregulation in NIPBL and cohesin mutant human cells. PLoS biology *7*, e1000119.

Lobanenkov, V.V., Nicolas, R.H., Adler, V.V., Paterson, H., Klenova, E.M., Polotskaja, A.V., and Goodwin, G.H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. Oncogene *5*, 1743-1753.

Losada, A., Hirano, M., and Hirano, T. (1998). Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. Genes & development *12*, 1986-1997.

Losada, A., Hirano, M., and Hirano, T. (2002). Cohesin release is required for sister chromatid resolution, but not for condensin-mediated compaction, at the onset of mitosis. Genes & development *16*, 3004-3016.

Losada, A., and Hirano, T. (2001). Intermolecular DNA interactions stimulated by the cohesin complex in vitro. Current Biology *11*, 268-272.

Losada, A., Yokochi, T., and Hirano, T. (2005). Functional contribution of Pds5 to cohesin-mediated cohesion in human cells and Xenopus egg extracts. Journal of cell science *118*, 2133-2141.

Losada, A., Yokochi, T., Kobayashi, R., and Hirano, T. (2000). Identification and characterization of SA/Scc3p subunits in the Xenopus and human cohesin complexes. The Journal of Cell Biology *150*, 405-416.

Löwe, J., Cordell, S.C., and van den Ent, F. (2001). Crystal structure of the SMC head domain: an ABC ATPase with 900 residues antiparallel coiled-coil inserted1. Journal of Molecular Biology *306*, 25-35.

Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R.*, et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012-1025.

Mahy, N.L., Perry, P.E., and Bickmore, W.A. (2002). Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. J Cell Biol *159*, 753-763.

Mannini, L., Cucco, F., Quarantotti, V., Krantz, I.D., and Musio, A. (2013). Mutation Spectrum and Genotype–Phenotype Correlation in Cornelia de Lange Syndrome. Human mutation *34*, 10.1002/humu.22430.

Martini, E., Roche, D.M., Marheineke, K., Verreault, A., and Almouzni, G. (1998). Recruitment of phosphorylated chromatin assembly factor 1 to chromatin after UV irradiation of human cells. J Cell Biol *143*, 563-575.

McGuinness, B.E., Hirota, T., Kudo, N.R., Peters, J.M., and Nasmyth, K. (2005). Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells. PLoS biology *3*, e86.

Melby, T.E., Ciampaglio, C.N., Briscoe, G., and Erickson, H.P. (1998). The Symmetrical Structure of Structural Maintenance of Chromosomes (SMC) and MukB Proteins: Long, Antiparallel Coiled Coils, Folded at a Flexible Hinge. The Journal of Cell Biology *142*, 1595-1604.

Merkenschlager, M., and Odom, D.T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. Cell *152*, 1285-1297.

Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: Chromosomal Proteins that Prevent Premature Separation of Sister Chromatids. Cell *91*, 35-45.

Misulovin, Z., Schwartz, Y.B., Li, X.Y., Kahn, T.G., Gause, M., MacArthur, S., Fay, J.C., Eisen, M.B., Pirrotta, V., Biggin, M.D.*, et al.* (2008). Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. Chromosoma *117*, 89-102.

Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.M., Taneja, N., Folco, H.D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J.*, et al.* (2014). Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. Nature *516*, 432-435.

Monahan, K., Rudnick, N.D., Kehayova, P.D., Pauli, F., Newberry, K.M., Myers, R.M., and Maniatis, T. (2012). Role of CCCTC binding factor (CTCF) and cohesin in the

generation of single-cell diversity of protocadherin-alpha gene expression. Proceedings of the National Academy of Sciences of the United States of America *109*, 9125-9130.

Moore, C.B., Guthrie, E.H., Huang, M.T., and Taxman, D.J. (2010). Short hairpin RNA (shRNA): design, delivery, and assessment of gene knockdown. Methods Mol Biol *629*, 141-158.

Murayama, Y., and Uhlmann, F. (2014). Biochemical reconstitution of topological DNA binding by the cohesin ring. Nature *505*, 367-371.

Murray, A.W., and Szostak, J.W. (1985). Chromosome segregation in mitosis and meiosis. Annual review of cell biology *1*, 289-315.

Musio, A., Selicorni, A., Focarelli, M.L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P., and Larizza, L. (2006). X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat Genet *38*, 528-530.

Nakahashi, H., Kwon, K.R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A.*, et al.* (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell reports *3*, 1678-1689.

Nakajima, M., Kumada, K., Hatakeyama, K., Noda, T., Peters, J.M., and Hirota, T. (2007). The complete removal of cohesin from chromosome arms depends on separase. Journal of cell science *120*, 4188-4196.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). Transcription. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017-1021.

Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. Annual review of genetics *43*, 525-558.

Nativio, R., Wendt, K.S., Ito, Y., Huddleston, J.E., Uribe-Lewis, S., Woodfine, K., Krueger, C., Reik, W., Peters, J.M., and Murrell, A. (2009). Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. PLoS Genet *5*, e1000739.

Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A.A., Mechtler, K.*, et al.* (2010). Sororin mediates sister chromatid cohesion by antagonizing Wapl. Cell *143*, 737-749.

Nitzsche, A., Paszkowski-Rogacz, M., Matarese, F., Janssen-Megens, E.M., Hubner, N.C., Schulz, H., de Vries, I., Ding, L., Huebner, N., Mann, M.*, et al.* (2011). RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. PLoS One *6*, e19470.

Nolen, L.D., Boyle, S., Ansari, M., Pritchard, E., and Bickmore, W.A. (2013). Regional chromatin decompaction in Cornelia de Lange syndrome associated with NIPBL disruption can be uncoupled from cohesin and CTCF. Human molecular genetics *22*, 4180-4193.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J.*, et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-385.

Orlando, D.A., Chen, M.W., Brown, V.E., Solanki, S., Choi, Y.J., Olson, E.R., Fritz, C.C., Bradner, J.E., and Guenther, M.G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell reports *9*, 1163-1170.

Ouyang, Z., Zheng, G., Song, J., Borek, D.M., Otwinowski, Z., Brautigam, C.A., Tomchick, D.R., Rankin, S., and Yu, H. (2013). Structure of the human cohesin inhibitor Wapl. Proc Natl Acad Sci U S A *110*, 11355-11360.

Panizza, S., Tanaka, T., Hochwagen, A., Eisenhaber, F., and Nasmyth, K. (2000). Pds5 cooperates with cohesin in maintaining sister chromatid cohesion. Current Biology *10*, 1557-1564.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T.*, et al.* (2008). Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. Cell *132*, 422-433.

Pauli, A., Althoff, F., Oliveira, R.A., Heidmann, S., Schuldiner, O., Lehner, C.F., Dickson, B.J., and Nasmyth, K. (2008). Cell-type-specific TEV protease cleavage reveals cohesin functions in Drosophila neurons. Developmental cell *14*, 239-251.

Peters, J.M., Tedeschi, A., and Schmitz, J. (2008). The cohesin complex and its roles in chromosome biology. Genes & development *22*, 3089-3114.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y.*, et al.* (2013). Architectural protein

subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281-1295.

Phillips, J.E., Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 194-211.

Poorey, K., Viswanathan, R., Carver, M.N., Karpova, T.S., Cirimotich, S.M., McNally, J.G., Bekiranov, S., and Auble, D.T. (2013). Measuring chromatin interaction dynamics on the second time scale at single-copy genes. Science *342*, 369-372.

Rankin, S., Ayad, N.G., and Kirschner, M.W. (2005). Sororin, a substrate of the anaphase-promoting complex, is required for sister chromatid cohesion in vertebrates. Mol Cell *18*, 185-200.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S.*, et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.

Remeseiro, S., Cuadrado, A., Carretero, M., Martinez, P., Drosopoulos, W.C., Canamero, M., Schildkraut, C.L., Blasco, M.A., and Losada, A. (2012a). Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. The EMBO journal *31*, 2076-2089.

Remeseiro, S., Cuadrado, A., Gomez-Lopez, G., Pisano, D.G., and Losada, A. (2012b). A unique role of cohesin-SA1 in gene regulation and development. The EMBO journal *31*, 2090-2102.

Rhee, H.S., and Pugh, B.F. (2012). ChIP-exo: A Method to Identify Genomic Location of DNA-binding proteins at Near Single Nucleotide Accuracy. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] *0 21*, 10.1002/0471142727.mb0471142124s0471142100.

Rhodes, J.M., Bentley, F.K., Print, C.G., Dorsett, D., Misulovin, Z., Dickinson, E.J., Crosier, K.E., Crosier, P.S., and Horsfield, J.A. (2010). Positive regulation of c-Myc by cohesin is direct, and evolutionarily conserved. Dev Biol *344*, 637-649.

Riedel, C.G., Katis, V.L., Katou, Y., Mori, S., Itoh, T., Helmhart, W., Galova, M., Petronczki, M., Gregan, J., Cetin, B.*, et al.* (2006). Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I. Nature *441*, 53-61.

Roig, M.B., Lowe, J., Chan, K.L., Beckouet, F., Metson, J., and Nasmyth, K. (2014). Structure and function of cohesin's Scc3/SA regulatory subunit. FEBS Lett *588*, 3692-3702.

Rolef Ben-Shahar, T., Heeger, S., Lehane, C., East, P., Flynn, H., Skehel, M., and Uhlmann, F. (2008). Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. Science *321*, 563-566.

Rollins, R.A., Korom, M., Aulner, N., Martens, A., and Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. Mol Cell Biol *24*, 3100-3111.

Rollins, R.A., Morcillo, P., and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics *152*, 577-593.

Rowland, B.D., Roig, M.B., Nishino, T., Kurze, A., Uluocak, P., Mishra, A., Beckouet, F., Underwood, P., Metson, J., Imre, R.*, et al.* (2009). Building sister chromatid cohesion: smc3 acetylation counteracts an antiestablishment activity. Mol Cell *33*, 763-774.

Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., and Krumm, A. (2008). CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A *105*, 8309-8314.

Salic, A., Waters, J.C., and Mitchison, T.J. (2004). Vertebrate shugoshin links sister centromere cohesion and kinetochore microtubule stability in mitosis. Cell *118*, 567-578.

Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P., and Odom, D.T. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res *20*, 578-588.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Goncalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell *148*, 335-348.

Schmitz, J., Watrin, E., Lenart, P., Mechtler, K., and Peters, J.M. (2007). Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. Current biology : CB *17*, 630-636.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S.*, et al.* (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nature genetics *42*, 53-61.

Schule, B., Oviedo, A., Johnston, K., Pai, S., and Francke, U. (2005). Inactivating mutations in ESCO2 cause SC phocomelia and Roberts syndrome: no phenotype-genotype correlation. American journal of human genetics *77*, 1117-1128.

Schwartz, Y.B., Linder-Basso, D., Kharchenko, P.V., Tolstorukov, M.Y., Kim, M., Li, H.B., Gorchakov, A.A., Minoda, A., Shanower, G., Alekseyenko, A.A.*, et al.* (2012). Nature and function of insulator protein binding sites in the Drosophila genome. Genome Res *22*, 2188-2198.

Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., and Gaul, U. (2008). Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature *451*, 535-540.

Seitan, V.C., Faure, A.J., Zhan, Y., McCord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G.*, et al.* (2013). Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. Genome Res *23*, 2066-2077.

Seitan, V.C., Hao, B., Tachibana-Konwalski, K., Lavagnolli, T., Mira-Bontenbal, H., Brown, K.E., Teng, G., Carroll, T., Terry, A., Horan, K.*, et al.* (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. Nature *476*, 467-471.

Sexton, T., and Cavalli, G. (2013). The 3D genome shapes up for pluripotency. Cell stem cell *13*, 3-4.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.

Shen, W., Wang, D., Ye, B., Shi, M., Zhang, Y., and Zhao, Z. (2015). A possible role of Drosophila CTCF in mitotic bookmarking and maintaining chromatin domains during the cell cycle. Biological research *48*, 27.

Silva, J., and Smith, A. (2008). Capturing pluripotency. Cell *132*, 532-536.

Skibbens, R.V., Corson, L.B., Koshland, D., and Hieter, P. (1999). Ctf7p is essential for sister chromatid cohesion and links mitotic chromosome structure to the DNA replication machinery. Genes Dev *13*, 307-319.

Sofueva, S., and Hadjur, S. (2012). Cohesin-mediated chromatin interactions--into the third dimension of gene regulation. Briefings in functional genomics *11*, 205-216.

Sofueva, S., Yaffe, E., Chan, W.C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. Embo J *32*, 3119-3129.

Solomon, D.A., Kim, T., Diaz-Martinez, L.A., Fair, J., Elkahloun, A.G., Harris, B.T., Toretsky, J.A., Rosenberg, S.A., Shukla, N., Ladanyi, M.*, et al.* (2011). Mutational inactivation of STAG2 causes aneuploidy in human cancer. Science *333*, 1039-1043.

Spitz, F., Gonzalez, F., and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell *113*, 405-417.

Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev *20*, 2349-2354.

Stedman, W., Kang, H., Lin, S., Kissil, J.L., Bartolomei, M.S., and Lieberman, P.M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. The EMBO journal *27*, 654-666.

Strom, L., Lindroos, H.B., Shirahige, K., and Sjogren, C. (2004). Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair. Mol Cell *16*, 1003-1015.

Sumara, I., Vorlaufer, E., Gieffers, C., Peters, B.H., and Peters, J.M. (2000). Characterization of vertebrate cohesin complexes and their regulation in prophase. J Cell Biol *151*, 749-762.

Sumara, I., Vorlaufer, E., Stukenberg, P.T., Kelm, O., Redemann, N., Nigg, E.A., and Peters, J.-M. (2002). The Dissociation of Cohesin from Chromosomes in Prophase Is Regulated by Polo-like Kinase. Molecular Cell *9*, 515-525.

Sun, Y., Kucej, M., Fan, H.-Y., Yu, H., Sun, Q.-Y., and Zou, H. (2009). Separase Is Recruited to Mitotic Chromosomes to Dissolve Sister Chromatid Cohesion in a DNA-Dependent Manner. Cell *137*, 123-132.

Sutani, T., Kawaguchi, T., Kanno, R., Itoh, T., and Shirahige, K. (2009). Budding Yeast Wpl1(Rad61)-Pds5 Complex Counteracts Sister Chromatid Cohesion-Establishing Reaction. Current biology : CB *19*, 492-497.

Tachibana-Konwalski, K., Godwin, J., van der Weyden, L., Champion, L., Kudo, N.R., Adams, D.J., and Nasmyth, K. (2010). Rec8-containing cohesin maintains bivalents without turnover during the growing phase of mouse oocytes. Genes Dev *24*, 2505-2516.

Tanaka, K., Hao, Z., Kai, M., and Okayama, H. (2001). Establishment and maintenance of sister chromatid cohesion in fission yeast by a unique mechanism. The EMBO journal *20*, 5779-5790.

Tanaka, K., Yonekawa, T., Kawasaki, Y., Kai, M., Furuya, K., Iwasaki, M., Murakami, H., Yanagida, M., and Okayama, H. (2000). Fission yeast Eso1p is required for establishing sister chromatid cohesion during S phase. Mol Cell Biol *20*, 3459-3469.

Tang, Z., Shu, H., Qi, W., Mahmood, N.A., Mumby, M.C., and Yu, H. (2006). PP2A is required for centromeric localization of Sgo1 and proper chromosome segregation. Developmental cell *10*, 575-585.

Tedeschi, A., Wutz, G., Huet, S., Jaritz, M., Wuensche, A., Schirghuber, E., Davidson, I.F., Tang, W., Cisneros, D.A., Bhaskara, V*., et al.* (2013). Wapl is an essential regulator of chromatin structure and chromosome segregation. Nature *501*, 564-568.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell *10*, 1453-1465.

Tonkin, E.T., Wang, T.J., Lisgo, S., Bamshad, M.J., and Strachan, T. (2004). NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. Nat Genet *36*, 636-641.

Toth, A., Ciosk, R., Uhlmann, F., Galova, M., Schleiffer, A., and Nasmyth, K. (1999). Yeast cohesin complex requires a conserved protein, Eco1p(Ctf7), to establish cohesion between sister chromatids during DNA replication. Genes Dev *13*, 320-333.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Uhlmann, F., Lottspeich, F., and Nasmyth, K. (1999). Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. Nature *400*, 37-42.

Uhlmann, F., and Nasmyth, K. (1998). Cohesion between sister chromatids must be established during DNA replication. Current biology : CB *8*, 1095-1101.

Uhlmann, F., Wernic, D., Poupart, M.A., Koonin, E.V., and Nasmyth, K. (2000). Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. Cell *103*, 375-386.

Unal, E., Heidinger-Pauli, J.M., Kim, W., Guacci, V., Onn, I., Gygi, S.P., and Koshland, D.E. (2008). A molecular determinant for the establishment of sister chromatid cohesion. Science *321*, 566-569.

Van Bortle, K., Nichols, M.H., Li, L., Ong, C.T., Takenaka, N., Qin, Z.S., and Corces, V.G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. Genome biology *15*, R82.

van de Werken, H.J.G., Landan, G., Holwerda, S.J.B., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Oz, Y., Bouwman, B.A.M*., et al.* (2012).

Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat Meth *9*, 969-972.

Vass, S., Cotterill, S., Valdeolmillos, A.M., Barbero, J.L., Lin, E., Warren, W.D., and Heck, M.M. (2003). Depletion of Drad21/Scc1 in Drosophila cells leads to instability of the cohesin complex and disruption of mitotic progression. Current biology : CB *13*, 208-218.

Vega, H., Waisfisz, Q., Gordillo, M., Sakai, N., Yanagihara, I., Yamada, M., van Gosliga, D., Kayserili, H., Xu, C., Ozono, K.*, et al.* (2005). Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nat Genet *37*, 468-470.

Verni, F., Gandhi, R., Goldberg, M.L., and Gatti, M. (2000). Genetic and molecular analysis of wings apart-like (wapl), a gene controlling heterochromatin organization in Drosophila melanogaster. Genetics *154*, 1693-1710.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell reports *10*, 1297-1309.

Vrouwe, M.G., Elghalbzouri-Maghrani, E., Meijers, M., Schouten, P., Godthelp, B.C., Bhuiyan, Z.A., Redeker, E.J., Mannens, M.M., Mullenders, L.H., Pastink, A.*, et al.* (2007). Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: evidence for impaired recombinational repair. Human molecular genetics *16*, 1478-1487.

Waizenegger, I.C., Hauf, S., Meinke, A., and Peters, J.M. (2000). Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. Cell *103*, 399-410.

Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R.*, et al.* (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res *22*, 1680-1688.

Watrin, E., Schleiffer, A., Tanaka, K., Eisenhaber, F., Nasmyth, K., and Peters, J.-M. (2006). Human Scc4 Is Required for Cohesin Binding to Chromatin, Sister-Chromatid Cohesion, and Mitotic Progression. Current Biology *16*, 863-874.

Webster, M., Witkin, K.L., and Cohen-Fix, O. (2009). Sizing up the nucleus: nuclear shape, size and nuclear-envelope assembly. J Cell Sci *122*, 1477-1486.

Weitzer, S., Lehane, C., and Uhlmann, F. (2003). A model for ATP hydrolysis-dependent binding of cohesin to DNA. Curr Biol *13*, 1930-1940.

Wendt, K.S., and Peters, J.M. (2009). How cohesin and CTCF cooperate in regulating gene expression. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology *17*, 201-214.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T.*, et al.* (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796-801.

Williamson, I., Eskeland, R., Lettice, L.A., Hill, A.E., Boyle, S., Grimes, G.R., Hill, R.E., and Bickmore, W.A. (2012). Anterior-posterior differences in HoxD chromatin topology in limb development. Development *139*, 3157-3167.

Wood, A.J., Severson, A.F., and Meyer, B.J. (2010). Condensin and cohesin complexity: the expanding repertoire of functions. Nature reviews Genetics *11*, 391-404.

Xiao, T., Wallace, J., and Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. Mol Cell Biol *31*, 2174-2183.

Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat Genet *43*, 1059-1065.

Yagi, T. (2008). Clustered protocadherin family. Dev Growth Differ *50 Suppl 1*, S131-140.

Zaidi, S.K., Young, D.W., Montecino, M.A., Lian, J.B., van Wijnen, A.J., Stein, J.L., and Stein, G.S. (2010). Mitotic bookmarking of genes: a novel dimension to epigenetic control. Nature reviews Genetics *11*, 583-589.

Zhang, B., Jain, S., Song, H., Fu, M., Heuckeroth, R.O., Erlich, J.M., Jay, P.Y., and Milbrandt, J. (2007). Mice lacking sister chromatid cohesion protein PDS5B exhibit

developmental abnormalities reminiscent of Cornelia de Lange syndrome. Development *134*, 3191-3201.

Zhang, J., Shi, X., Li, Y., Kim, B.-J., Jia, J., Huang, Z., Yang, T., Fu, X., Jung, S.Y., Wang, Y.*, et al.* (2008). Acetylation of Smc3 by Eco1 Is Required for S Phase Sister Chromatid Cohesion in Both Human and Yeast. Mol Cell *31*, 143-151.

Zhang, N., Jiang, Y., Mao, Q., Demeler, B., Tao, Y.J., and Pati, D. (2013). Characterization of the interaction between the cohesin subunits Rad21 and SA1/2. PLoS One *8*, e69458.

Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van, I.W.F.*, et al.* (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci U S A *111*, 996-1001.